JMIR Formative Research

Contents

Original Papers

XSL•FO
RenderX

XSL•FO

RenderX

Early Reports

Viewpoint

Notes From the Field

Corrigenda and Addenda

Original Paper

Heart Rate Measurement Accuracy of Fitbit Charge 4 and Samsung Galaxy Watch Active2: Device Evaluation Study

Michael Nissen[1], MSc; Syrine Slim[1], BSc; Katharina Jäger[1], MSc; Madeleine Flaucher[1], MSc; Hanna Huebner[2], Dr rer nat; Nina Danzberger[2], MSc; Peter A Fasching[2], Prof Dr med; Matthias W Beckmann[2], Prof Dr med; Stefan Gradl[1], Dr Ing; Bjoern M Eskofier[1], Prof Dr

[1]Department Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany
[2]Department of Gynecology and Obstetrics, Erlangen University Hospital, Erlangen, Germany

Corresponding Author:
Michael Nissen, MSc
Department Artificial Intelligence in Biomedical Engineering
Friedrich-Alexander University Erlangen-Nürnberg
Schlossplatz 4
Erlangen, 91054
Germany
Phone: 49 913185 28990
Email: michael.nissen@fau.de

Abstract

Background: Fitness trackers and smart watches are frequently used to collect data in longitudinal medical studies. They allow continuous recording in real-life settings, potentially revealing previously uncaptured variabilities of biophysiological parameters and diseases. Adequate device accuracy is a prerequisite for meaningful research.

Objective: This study aims to assess the heart rate recording accuracy in two previously unvalidated devices: Fitbit Charge 4 and Samsung Galaxy Watch Active2.

Methods: Participants performed a study protocol comprising 5 resting and sedentary, 2 low-intensity, and 3 high-intensity exercise phases, lasting an average of 19 minutes 27 seconds. Participants wore two wearables simultaneously during all activities: Fitbit Charge 4 and Samsung Galaxy Watch Active2. Reference heart rate data were recorded using a medically certified Holter electrocardiogram. The data of the reference and evaluated devices were synchronized and compared at 1-second intervals. The mean, mean absolute error, mean absolute percentage error, Lin concordance correlation coefficient, Pearson correlation coefficient, and Bland-Altman plots were analyzed.

Results: A total of 23 healthy adults (mean age 24.2, SD 4.6 years) participated in our study. Overall, and across all activities, the Fitbit Charge 4 slightly underestimated the heart rate, whereas the Samsung Galaxy Watch Active2 overestimated it (−1.66 beats per minute [bpm]/3.84 bpm). The Fitbit Charge 4 achieved a lower mean absolute error during resting and sedentary activities (seated rest: 7.8 vs 9.4; typing: 8.1 vs 11.6; laying down [left]: 7.2 vs 9.4; laying down [back]: 6.0 vs 8.6; and walking slowly: 6.8 vs 7.7 bpm), whereas the Samsung Galaxy Watch Active2 performed better during and after low- and high-intensity activities (standing up: 12.3 vs 9.0; walking fast: 6.1 vs 5.8; stairs: 8.8 vs 6.9; squats: 15.7 vs 6.1; resting: 9.6 vs 5.6 bpm).

Conclusions: Device accuracy varied with activity. Overall, both devices achieved a mean absolute percentage error of just <10%. Thus, they were considered to produce valid results based on the limits established by previous work in the field. Neither device reached sufficient accuracy during seated rest or keyboard typing. Thus, both devices may be eligible for use in respective studies; however, researchers should consider their individual study requirements.

KEYWORDS

XSL•FO
RenderX

Introduction

Background

Wearables such as smart watches and fitness trackers enable data recording in real-life settings, where biomedical signals cannot be easily captured with conventional or clinical devices. They can provide unobtrusive, economic, high-resolution, longitudinal recording capabilities for various signals, including accelerometer and photoplethysmogram (PPG) data [1,2]. This makes them particularly interesting for use in longitudinal medical studies and biomedical research. Observational studies especially benefit from the unobtrusive and longitudinal recording characteristics of fitness trackers and wearables across a variety of medical disciplines [3-9]. In addition to clinical studies, fitness trackers are also usable in other applications. These include activity feedback, activity promotion, weight management, disease monitoring, disease diagnostics, stress and sleep monitoring, and health care surveillance [10-15].

An important prerequisite for the use of wearables in studies and connected applications is adequate accuracy and, thus, data quality. Without sufficient validation, the reliability of the recorded data is unknown, which is the case for many modern end consumer devices. Thus, stringent upfront validation is not only a necessity for meaningful research but also prospective applications.

Related Work

The Fitbit Charge HR series (Fitbit) is one of the most frequently validated devices. The study by Lee [16] reviewed the first device generation with 10 college students under free-living conditions. Each participant was asked to conduct normal day activities for 8 hours, and the heart rate (HR) was recorded and evaluated every minute against a Polar HR chest strap monitor. They concluded that the device was not accurate, with a mean absolute percentage error (MAPE) of 9.17% (SD 10.9%) when worn on the nondominant hand. Brazendale et al [17] evaluated the HR measurements of 39 children. The evaluation was performed on a per-minute basis, and the MAPE was reported as 6.9%. Thus, the authors concluded that wearable fitness trackers provide HR measurements comparable with a criterion field–based measure.

The data of 50 intensive care unit patients monitored over 24 hours were used for evaluation by Kroll et al [15]. They recorded HR values every 5 minutes and identified a median difference of 1 beats per minute (bpm) between the derived HR of the fitness tracker and the electrocardiogram (ECG)–derived HR.

A higher comparison frequency was chosen by Jo et al [18]. By measuring and comparing the HR every second, 24 participants completed a 77-minute protocol comprising several activities, including cycling, walking, jogging, running, and other sports exercises. A 12-lead ECG served as the criterion device. The authors reported a mean bias of −8.8 bpm and concluded that the device by Fitbit does not satisfy the validity criteria, particularly during higher exercise intensities.

The second release of the Fitbit Charge HR series was evaluated by Reddy et al [19], Thomson et al [20], and Benedetto et al [21], yielding different results on the device accuracy. To the best of our knowledge, the only validation study for Fitbit Charge 3 was performed by Muggeridge et al [22], who stated that the Fitbit Charge 3 performed well only during resting and walking-like conditions but otherwise assessed the accuracy to be overall poor.

To the best of our knowledge, no validation studies on the Samsung Galaxy Watch Active series exist as of today. However, other Samsung smart watches have been validated in the past. The measurements of the Samsung Gear S were investigated by Wallen et al [23], with 22 participants in rest, walking, running, and cycling. Out of a total of 4 devices, Samsung Gear S demonstrated the greatest variability in HR measurements. Shcherbina et al [24] examined the Samsung Gear S2 among 6 other devices with 60 participants from diverse backgrounds, performing a range of activities, including sitting, walking, running, and cycling. Of the validated devices in the study, Samsung Gear S2 showed the highest overall error, particularly during sitting. In another study by El-Amrawy and Nounou [25], the device also showed the lowest accuracy compared with 17 other devices and a clinical pulse oximeter as a criterion device.

Objective

The validation study presented here was conducted as an initial groundwork for a large-scale observational study in obstetrics. As a pilot study, we aim to assess the performance of the selected devices in a healthy population to initially determine eligibility for longitudinal medical studies in general. We were particularly interested in the performance and accuracy of HR measurements, which are analyzed in detail in the following sections. This work is the first to validate the Fitbit Charge 4 and Samsung Galaxy Watch Active2 (Samsung Group).

Methods

Overview

Details on the participants, experimental procedure, used devices, validation metrics, processing, and evaluation are outlined in the following sections. Where applicable and possible, we adhered to several common grounds, guidelines, and best practices for wearable HR validation, which have been published in the more recent past [2,12,26].

Ethics Approval

The study was approved by the ethics committee of Friedrich–Alexander Universität Erlangen-Nürnberg (106_13 B). The participants provided informed consent to participate.

Recruitment

Recruitment was conducted via mailing lists and direct contact. We were unable to perform a power calculation for sample size estimation as the selected devices have not been investigated in the past, and thus, no information on effect sizes or variances was available. Consequently, we aimed at a sample size of approximately 20 to 25 participants, which is in line with previous HR evaluation studies [18,19,22,27,28]. Exclusion criteria was a major underlying medical condition affecting the participants' physical capability or increasing the risk of injury. Assessment was conducted using the Physical Activity

XSL•FO

RenderX

Readiness Questionnaire [29]. Ultimately, 23 participants were recruited.

Devices and Gold Standard

We aimed to investigate the accuracy of Fitbit Charge 4 and Samsung Galaxy Watch Active2. Fitbit Charge 4 was released in March 2020. According to the manufacturer, it has a battery runtime of up to 7 days, which makes it particularly interesting for longitudinal studies [30].

The Galaxy Watch Active2 is a smart watch using Tizen OS as the operating system. Its PPG sensor uses 8 photodiodes [31]. The smart watch is available in several sizes and editions; our study used a 40 mm–sized version without long-term evolution. Furthermore, the device is able to record the ECGs. It is possible to derive blood pressure measurements through the PPG sensor; an upfront validation with a blood pressure cuff is required beforehand, and it is recommended to repeat this validation every 4 weeks [32,33]. Tizen OS is extendable, and a documentation of several available application programming interfaces (APIs) to create custom applications and interact with the device is available on the web [34]. This includes functions for accessing nearly all the built-in sensors. The available functions are not limited to the reading of HR and RR intervals but also allow raw data access to nearly all built-in sensors, particularly the PPG sensor. These features make the device interesting for medical studies, as their own algorithms for data processing can be used.

A Mind Media NeXus-10 MKI (Mind Media BV) was used as the gold standard. The ECG Holter device is a certified medical device of class 2a (EU). Data are transferred in real time via Bluetooth to a computer running a manufacturer-supplied software called Biotrace+ (Mind Media) [35], which displays and allows the export of HR, heart rate variability, and ECG data.

Study Procedure

The study was conducted in an indoor laboratory environment on 7 different days between July 30, 2020, and September 21, 2020. As the data recording was conducted during summer without air conditioning, ambient temperatures were comparably high for Northern Bavaria, causing sweaty skin surfaces in some cases. This can induce additional noise, electrode loss, or affect the PPG signal measurement of wearable devices.

After receiving information on the study procedure and aims, participants filled out the activity readiness questionnaire and respective consent forms. Participants were then supplied with the 2 wearable devices and were asked to place 1 device on each arm, ensuring that the sensor was in good contact with the skin and that the devices were fitted comfortably on the arms. The study adviser determined which device should be placed onto which arm, and devices were equally placed on the left or right arm across all participants. As the Fitbit app has a setting to determine whether the device is placed on the dominant or nondominant arm, this setting was configured accordingly by the study adviser based on the participant information. No such setting exists for the Galaxy Watch Active2. Subsequently, the Mind Media NeXus-10 MKI's electrodes were placed in a lead two position. To reduce noise, electrodes were placed on the torso, not the extremities. As the Holter device was equipped with a handbag-like body strap, it was hung over the shoulder of participants to increase freedom of movement. This positioning method was supported by the manufacturer. Recording started at least 30 seconds after placement of the electrodes, ensuring sufficient time for adaption for both wearables and the ECG algorithms.

The participants conducted an experimental protocol covering 10 subsequent tasks. Each task lasted between 1 and 2 minutes. The protocol anticipated a total length of 15 minutes. The chosen activities originate from activity recommendations for women with pregnancies, who are the prospective target group in our anticipated larger study. Participants were asked to conduct activities at their own pace to resemble activities as they would be conducted *free living* by the target group. With transitions between the individual study protocol phases, the recordings had an average duration of 19:03 minutes. We tried to minimize transition or relaxation phases between activities to ensure that the respective HR levels were similar between adjunct activities. If minor slack times (usually <10 seconds) occurred between activities (eg, because of instructions by the study adviser or a move of position between activities), these slack times were not included in the individual activity analysis. Our overall goal was to initially start with resting and sedentary activities (seated rest, typing, laying down [left], and laying down [back]), then continually increase HR using low intensity (standing up and walking at a slow pace) and high intensity (walking at brisk pace, climbing stairs, and squat work out) activities. The full list of activities and tasks is presented in Table 1.

Table 1. List of conducted activities.

Activities	Duration (minute)	Description
Seated rest	2	• Sit comfortably on a chair while breathing normally, without any physical movements • Considered resting or baseline condition
Keyboard typing	1.5	• Type a neutral text on a computer keyboard provided by the study advisor • Aims to assess the effects of hand movement without general body movement
Laying on left side	1.5	• Lay on the left body side on a flat mattress
Laying on back	1.5	• Turn and lay down fully on the back
Standing up	1	• Stand up and maintain an upright position without movement
Walking at a slow pace	1	• Walk slowly and naturally around the laboratory at one's own pace
Walking at a brisk pace	2	• Increase the walking speed to the maximum walking speed without running at one's own pace
Climbing stairs	1.5	• Climb stairs up and down at one's own pace Simulation of a workout for a woman with pregnancy
Squat workout	1.5	• Conduct squats at one's own pace • Simulation of a workout for a woman with pregnancy
Seated rest	1.5	• Sit down directly after the workout, relax your breathing, and remain without motion. • Aims to assess drastic changes in heart rate from high activity to rest

Data Recording and Processing

Fitbit Charge 4

To increase the sampling frequency of the Fitbit Charge 4, the device was set to the training mode before the first study run. This produced a HR measurement every 1 to 5 seconds, thus resulting in a sampling frequency between 0.2 and 1 Hz. As the fitness tracker is linked to a user account in Fitbit's cloud, the data were accessed through the Fitbit Web API using representational state transfer queries and the Postman software. The API only provides access to HR measurements, and no PPG raw data or RR intervals are provided.

To compare data on a per-second basis, the HR values required upsampling. When not provided with a HR measurement every second, missing values were imputed using the next available HR value.

Samsung Galaxy Watch Active2

A custom application for Samsung's Tizen OS was developed. The *Human Activity Monitor API* was used to retrieve HR and RR intervals. All retrieved data were saved in JSON format to files and downloaded to a computer. The *Human Activity Monitor API* provides HR and RR interval data with a sampling rate of 25 Hz (the provided callback function to the *humanactivitymonitor.start* function is called every 40 milliseconds). However, the data are inconclusive: the HR changes more frequently than physiologically explainable; that is, the API provides up to 5 HR changes (from 86 to 85 to 86 to 85 to 84) within a time frame as small as 300 milliseconds. At the same time, the reported RR interval occasionally remains unchanged over periods >10 seconds. Thus, we decided to sample the HR and RR interval data at 1 Hz. A minority of the data was sampled at a higher frequency and manually downsampled to 1 Hz.

Mind Media NeXus-10 MKI

The criterion device recorded ECG data at a sampling frequency of 256 Hz. Furthermore, it contained internal peak detection algorithms, also providing derived HR and RR interval data at 32 Hz. As stated before, data were transferred via Bluetooth from the Holter device to a computer running Mind Media Biotrace+. The data were then exported from the Mind Media Biotrace+ software as a CSV file. Activity sections were recorded and annotated by the study adviser during the study execution in software running on a laptop computer.

Owing to the nature of the study protocol, some activities were prone to noise. Particularly during squats and stair climbing, ECGs were sometimes noisy, and the manufacturer-supplied software was apparently unable to correctly identify R peaks, resulting in erratic and evidently wrong HR and RR interval data. This was particularly true for squat and stair-climbing activities.

To cope with this issue, the raw criterion ECG was again processed in Python, using the ECG function from the BioSPPY library [36]. Subsequently, the R peaks were manually revised by a human annotator and corrected. We then used a self-developed function to extract the HR from the RR intervals.

Finally, the data were downsampled to 1 Hz. If >1 HR value occurred during 1 second (as the HR was >60 bpm), the respective values were averaged.

Data Exclusion

Although manual data processing was applied to ensure high data quality, some recorded criterion ECGs were too noisy and

unusable for comparison. Data were excluded if adequate criterion device recordings were unavailable but not if the measured data of the examined devices were evidently incorrect, as this situation could also appear in real-life use.

As a result, data of 3 participants (IDs 6, 8, and 23) had to be completely excluded. In addition, data of 2 participants were excluded for the *squat* and *walking stairs* activity (ID 5 and ID 7). After the *squat* activity, electrodes of 2 participants (ID 2 and ID 7) detached, and thus, no data were available. As stated before, a detailed overview of the conducted manual data correction and excluded activities of individual participants is provided in Table 2.

Table 2. Data exclusion and annotation.

Participant	Seated rest	Keyboard typing	Laying on left side	Laying on back	Standing up	Walking slow	Walking brisk	Stairs	Squats	Seated rest
1	Original[a]	Original	Annotated[b]	Original	Annotated	Original	Original	Original	Original	Original
2	Original	Original	Original	Original	Original	Original	Original	Original	Original	Excluded[c]
3	Original	Original	Annotated	Annotated	Annotated	Original	Annotated	Annotated	Original	Original
4	Original	Annotated	Annotated	Annotated	Original	Annotated	Annotated	Annotated	Annotated	Original
5	Original	Original	Annotated	Original	Original	Annotated	Annotated	Excluded	Excluded	Original
6	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded
7	Original	Original	Original	Original	Original	Annotated	Annotated	Excluded	Excluded	Excluded
8	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded
9	Original	Original	Original	Original	Original	Original	Original	Original	Original	Original
10	Original	Original	Original	Original	Annotated	Original	Original	Annotated	Original	Original
11	Original	Original	Original	Annotated	Original	Original	Original	Original	Annotated	Original
12	Original	Original	Original	Original	Original	Original	Original	Original	Original	Original
13	Original	Original	Original	Original	Original	Original	Original	Original	Original	Original
14	Original	Original	Original	Original	Original	Original	Original	Original	Annotated	Original
15	Original	Original	Original	Original	Original	Original	Original	Original	Annotated	Original
16	Original	Original	Original	Original	Annotated	Original	Original	Original	Original	Original
17	Original	Original	Original	Original	Original	Original	Original	Original	Original	Original
18	Annotated	Original	Original	Original	Annotated	Original	Original	Original	Original	Original
19	Original	Original	Original	Original	Original	Original	Original	Original	Annotated	Original
20	Original	Original	Original	Original	Annotated	Original	Original	Original	Annotated	Original
21	Original	Original	Original	Original	Original	Original	Original	Original	Original	Original
22	Original	Original	Original	Original	Original	Original	Original	Original	Original	Original
23	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded

[a]Represents original data.

[b]Represents manually annotated data.

[c]Represents excluded activity participant combinations.

Data Synchronization

As the software of the validated fitness trackers is mostly closed source, their exact time measurement and determination mechanism are unknown. Furthermore, the on-device signal processing may cause additional delays. Therefore, we did not rely on exact time stamps for device synchronization but instead used another synchronization technique.

Synchronization of the signals was performed on the previously downsampled signals of all 3 devices (1 Hz, ie, 1 HR value per second). We conducted the synchronization between the individual validated devices and our HR reference by maximizing the Pearson correlation coefficient (PCC). The

measurements were then shifted by the respectively determined time delays. This provided very similar and, in many cases, equal results to a shift through cross-correlation but showed better visual and metric results in a minority of edge cases.

Statistical Analysis

All statistical analyses were conducted in Python (version 3.8.7) on a Windows 10 machine using Numpy 1.19.5 [37], Scipy 1.6.0 [38], Pandas 1.2.0 [39], and Pingouin 0.3.11 [40]. Raw data and respective scripts are available from the authors upon request.

Absolute error analysis was conducted using the mean absolute error (MAE) and MAPE as key metrics. We defined MAE as

the average absolute distance between the HR of the validated device and the criterion device. MAPE is the percentage difference between the reference and the respective device values. The limits of agreement and mean error (bias) were derived from Bland–Altman plots, which also visually aided in the interpretation of the results. Correlation analysis was performed using the Lin concordance correlation coefficient (CCC), as suggested by Sartor et al [12,41,42]. PCC was additionally reported for completeness but not analyzed.

Results

Participants

In total, 23 healthy individuals participated in the study (n=10, 43% women and n=13, 57% men). The demographics and details of the participants are shown in Table 3. Most participants were university students and staff members. Given the location of the university, Fitzpatrick skin type 2 was overrepresented (3×type 1, 15×type 2, 2×type 3, 1×type 4, and 2×type 5).

Table 3. Demographics and details of the participants.

Characteristics	Values, minimum	Values, maximum	Values, mean (SD)
Age (years)[a]	20	36	24.2 (4.6)
Height (cm)	156	193	175.7 (10.48)
Body weight (kg)[b]	50	88	71 (12.75)

[a]One participant did not provide his or her date of birth.

[b]One participant did not provide his or her body weight.

HR Measurement

The key results of this validation study are summarized in Table 4. In total and across the entire experiment duration (ie, all activities), both devices achieved very similar values for MAE, MAPE, and PCC. Although the Fitbit Charge 4 slightly underestimated the HR by −1.66 bpm (bias), the Samsung Galaxy Watch Active2 overestimated the HR by 3.84 bpm (bias).

In resting and sedentary activities (seated rest, typing, and laying down) and slow walking, the Fitbit Charge 4 achieved lower absolute and absolute percentage error rates. During standing up and all other physical activities, the Samsung Galaxy Watch Active2 outperformed the Fitbit Charge 4.

A particularly high bias (ie, mean difference) was observed by the Fitbit Charge 4 during standing up (−7.95 bpm) and squats (−12.52 bpm). The Samsung Galaxy Watch Active2's highest bias was measured during typing (8.63 bpm) and laying down on the left side (6.01 bpm).

The level of agreement of the Samsung Galaxy Watch Active2 is particularly broad during activities 1 to 5. The cause was a non- or excessive recorded HR in participant 20 during these activities, where the device recorded an average HR of 146, 148, 181, and 176 bpm. This HR trend is displayed in Figure 1. If this participant was excluded from the data analysis, the metrics drastically improved: MAE and MAPE were consistently lower than those of the Fitbit device for activities 1 to 5, and both widths of limits of agreement and bias were reduced significantly.

CCC was consistently higher in the Samsung Galaxy Watch Active2. Both devices achieved particularly low scores (<0.250) during typing and slow walking. The resting phase resulted in the highest individual activity of CCC in both devices.

Bland–Altman plots for both devices are shown in Figures 2 and 3. A large cluster of points in the top-right section of Figure 3 is particularly noticeable. These data points are a result of the previously mentioned mismeasurement of the Samsung Galaxy Watch Active2. If participant 20 is excluded from the data set, the respective cluster disappears from the plot.

Table 4. Validation metrics of heart rate measurements across different activities and devices.

Activity (metrics)	Mind Media NeXus-10 MKI	Fitbit Charge 4	Samsung Galaxy Watch Active2
Overall (0)			
Values (bpm[a]), mean (SD)	92.48 (22.34)	90.85 (18.75)	96.26 (22.55)
MAE[b] (bpm)	N/A[c]	8.589	8.13
MAPE[d]	N/A	9.74	9.419
CCC[e]	N/A	0.805	0.847
PCC[f]	N/A	0.839	0.85
Bias (bpm)	N/A	−1.66	3.84
LoA[g] (bpm)	N/A	−26.75 to 23.43	−28.1 to 35.78
Seated rest (1)			
Values (bpm), mean (SD)	75.65 (7.17)	78.75 (3.88)	79.89 (9.43)
MAE (bpm)	N/A	7.829	9.381
MAPE	N/A	11.801	12.013
CCC	N/A	0.203	0.508
PCC	N/A	0.257	0.556
Bias (bpm)	N/A	3.36	4.41
LoA (bpm)	N/A	−18.98 to 25.70	−40.5 to 49.33
Typing (2)			
Values (bpm), mean (SD)	78.53 (6.23)	79.44 (3.1)	87.32 (5.84)
MAE (bpm)	N/A	8.139	11.62
MAPE	N/A	11.24	14.815
CCC	N/A	0.057	0.207
PCC	N/A	0.094	0.211
Bias (bpm)	N/A	0.79	8.63
LoA (bpm)	N/A	−1.96 to 22.14	−33.18 to 50.43
Laying down (left; 3)			
Values (bpm), mean (SD)	73.57 (9.17)	75.25 (4.48)	79.74 (8.94)
MAE (bpm)	N/A	7.245	9.35
MAPE	N/A	9.938	11.901
CCC	N/A	0.382	0.624
PCC	N/A	0.507	0.663
Bias (bpm)	N/A	2.13	6.01
LoA (bpm)	N/A	−17.49 to 21.75	−31.6 to 43.62
Laying down (back; 4)			
Values (bpm), mean (SD)	68.49 (7.75)	68.51 (3.61)	73.52 (5.25)
MAE (bpm)	N/A	6.034	8.622
MAPE	N/A	9.062	11.242
CCC	N/A	0.249	0.554
PCC	N/A	0.358	0.609
Bias (bpm)	N/A	0.03	4.82
LoA (bpm)	N/A	−16.62 to 16.67	−33.7 to 43.34
Standing up (5)			

Activity (metrics)	Mind Media NeXus-10 MKI	Fitbit Charge 4	Samsung Galaxy Watch Active2
Values (bpm), mean (SD)	88.83 (11.76)	81.03 (6.69)	88.68 (9.14)
MAE (bpm)	N/A	12.25	8.976
MAPE	N/A	13.302	9.988
CCC	N/A	0.253	0.519
PCC	N/A	0.345	0.62
Bias (bpm)	N/A	−7.95	0.52
LoA (bpm)	N/A	−37.83 to 21.93	−30.19 to 31.23
Walking slow (6)			
Values (bpm), mean (SD)	86.24 (6.38)	87.3 (3.86)	90.99 (3.53)
MAE (bpm)	N/A	6.781	7.742
MAPE	N/A	8.118	9.046
CCC	N/A	0.15	0.18
PCC	N/A	0.188	0.24
Bias (bpm)	N/A	1.2	4.94
LoA (bpm)	N/A	17.77 to 20.16	−15.45 to 25.34
Walking fast (7)			
Values (bpm), mean (SD)	100.22 (6.64)	99.11 (4.42)	102.82 (4.92)
MAE (bpm)	N/A	6.094	5.829
MAPE	N/A	6.364	6.292
CCC	N/A	0.348	0.439
PCC	N/A	0.408	0.516
Bias (bpm)	N/A	−0.98	2.86
LoA (bpm)	N/A	−20.68 to 18.73	−16.8 to 22.51
Stairs (8)			
Values (bpm), mean (SD)	119.67 (13.83)	115.54 (9.45)	121.14 (10.19)
MAE (bpm)	N/A	8.811	6.879
MAPE	N/A	7.605	6.157
CCC	N/A	0.634	0.691
PCC	N/A	0.803	0.812
Bias (bpm)	N/A	−3.99	1.28
LoA (bpm)	N/A	−25.61 to 17.63	−19.14 to 21.7
Squats (9)			
Values (bpm), mean (SD)	129.05 (11.87)	116.6 (7.72)	130.26 (7.28)
MAE (bpm)	N/A	15.737	6.163
MAPE	N/A	11.976	5.51
CCC	N/A	0.29	0.61
PCC	N/A	0.335	0.668
Bias (bpm)	N/A	−12.52	1.18
LoA (bpm)	N/A	−50.46 to 25.42	−20.55 to 22.92
Resting (10)			
Values (bpm), mean (SD)	106.17 (16.79)	102.8 (11.25)	105.82 (14.12)
MAE (bpm)	N/A	9.612	5.618
MAPE	N/A	9.749	5.82

XSL•FO
RenderX

Activity (metrics)	Mind Media NeXus-10 MKI	Fitbit Charge 4	Samsung Galaxy Watch Active2
CCC	N/A	0.65	0.845
PCC	N/A	0.752	0.872
Bias (bpm)	N/A	−3.45	−0.38
LoA (bpm)	N/A	−28.52 to 21.61	−15.75 to 15.00

[a]bpm: beats per minute.

[b]MAE: mean absolute error.

[c]N/A: not applicable.

[d]MAPE: mean absolute percentage error.

[e]CCC: Lin concordance correlation coefficient.

[f]PCC: Pearson correlation coefficient.

[g]LoA: limits of agreement.

Figure 1. Heart rate measurement of participant 20. Samsung Galaxy Watch Active2 recorded no or excessive heart rate values during the first 5 activities.
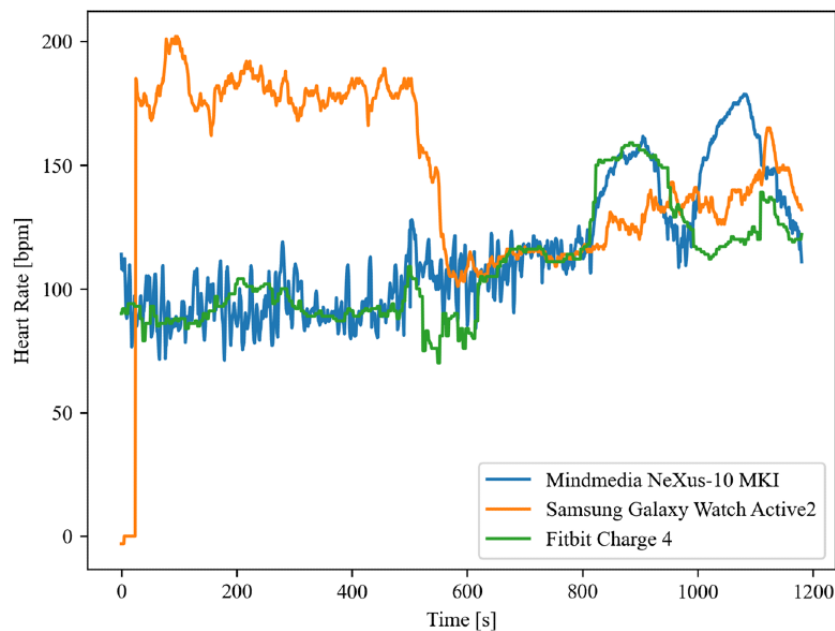
Figure 2. Bland-Altman plot for heart rate difference between Mind Media NeXus-10 MKI and Fitbit Charge 4 across all participants and activities.
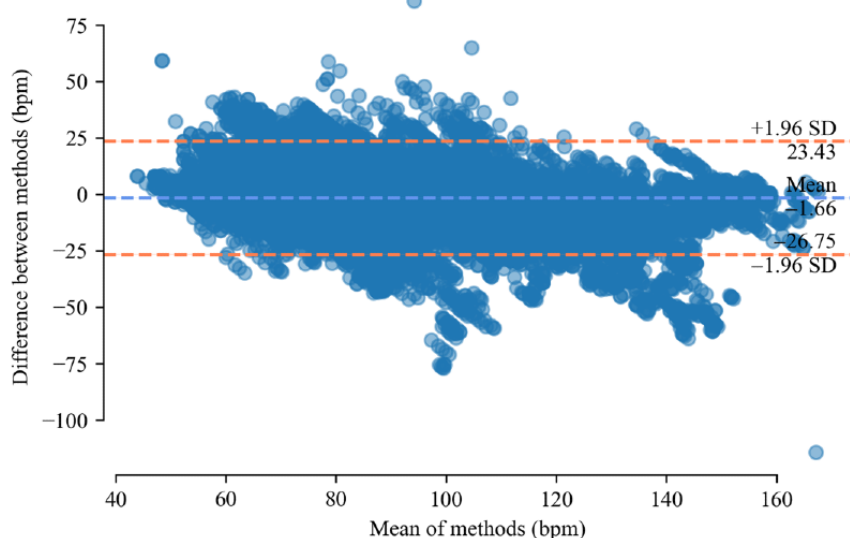
Figure 3. Bland-Altman plot for heart rate difference between Mind Media NeXus-10 MKI and Samsung Galaxy Watch Active2 across all participants and activities.
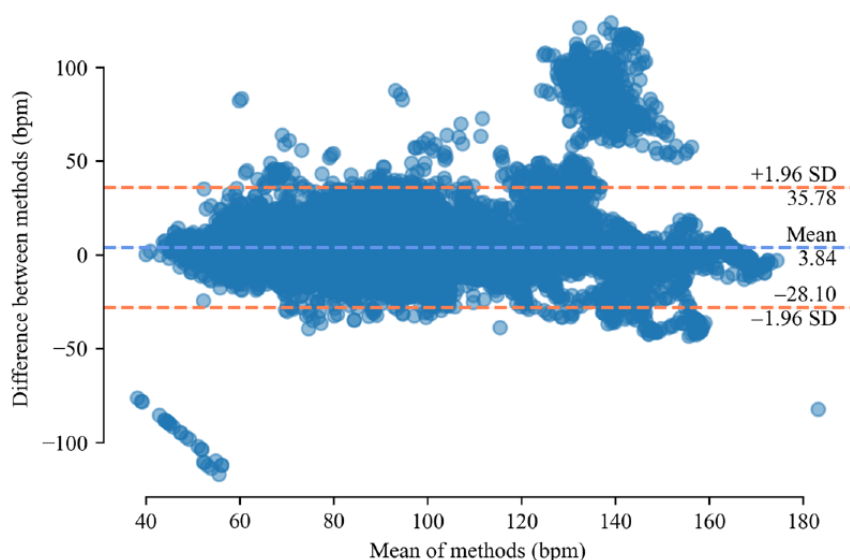
Discussion

Comparison With Previous Work

Our study aimed to evaluate 2 consumer wearable devices in healthy participants over a range of activities. The results from our study indicate that both devices achieved a MAPE <10%. Although no previous work exists for the Samsung Galaxy Watch Active2, our results are somewhat in line with previous validation trials for the Fitbit Charge series.

A previous evaluation of the Fitbit Charge 3 by Muggeridge et al [22] used a notably different experimental protocol, emphasizing strenuous activities (with a focus on treadmill running, sprinting, and cycling). The authors report an overall MAPE of 7.37 (as compared with 9.74 in our study) and note that the device underestimates the HR by −7 bpm (here, −1.66 bpm). Overall, the study states that the Fitbit device performs poorly during high-intensity activities and results in a higher error in that area. In our study, the Fitbit device's mean bias

was highest while climbing stairs and squatting, with a bias of −3.99 bpm and −12.52 bpm, respectively.

Reviewing studies on the Fitbit Charge 2, underestimations of the HR have been reported by several other studies [19,21,43]. The study by Baek et al [44] only reported this underestimation in the <100 bpm category and an overestimation of >120 bpm. With respect to the MAPE, the study by Reddy et al [19] reported a value of 11.33%, and the study by Nelson et al [43] reported a value of 5.96% across all activities. Our measured CCC of 0.805 across all activities is lower than the CCC of 0.906 reported by Nelson et al [43] in a 24-hour period.

Measurement Validity

Different validation definitions exist in the literature. Some prior studies have used an error rate of +5% to −5% as a limit, as it "approximates a widely accepted standard for statistical significance [...]" [24] and is "widely accepted" [19]. A limit of +10% to −10% is established by various organizations and

institutions and has been equally used by other validation studies [43]. The latter value is also proposed by previously mentioned validation guidelines [12] and thus, used for further reference in our work.

Similarly, different interpretations of correlation coefficients have been used in the literature. Owing to the large number of different definitions, ranging from a *weak* or *poor* interpretation starting between <0.2, <0.50, and <0.9 [23,43,45], we refrain from the use of an exact definition.

In our study and across all activities, both devices achieved a MAPE <10% and, per definition, produced valid results. With respect to individual activities, neither device produced valid results for seated rest and typing activities. Furthermore, the Fitbit Charge 4 did not record valid data for the standing up and squat activities, and the Samsung Galaxy Watch Active2 produced invalid results for laying down in either of the 2 evaluated positions.

Limitations

Participants and Demographic Structure

Our study mainly included healthy young participants aged between 20 and 36 years. Wearables may provide different validation results for older participants, particularly with respect to their skin properties and changes in the PPG curve. Furthermore, as most of our participants were local university students in middle Europe, Fitzpatrick skin types 1 to 3 were overrepresented in our study.

Selected Activities

The overall duration of individual activities was rather short, mostly because our aim was to set a low burden for study participation. Although the HR of all participants increased during the study duration (especially during the second half of the study), some participants may require a longer activity duration for optimal HR adaption. A shorter activity duration makes the collected data less meaningful and results in a lower number of recorded data points, thus decreasing statistical expressiveness.

As all activities were conducted consecutively and without breaks, splits between individual recorded activities always resulted in minor transitional phases. Some participants may react faster to the instructions of the study instructor than others. This leads to additional time slack between individual activities and may cause a slight metric profusion between the 2 subsequent activity metrics.

Laboratory Conditions and Environmental Factors

Although we aimed to replicate real-life activities as much as possible, our study was still conducted in a laboratory setting. Real-life use patterns may differ from those in our study and, as such, may have an impact on the accuracy of the investigated devices. Furthermore, our study was mostly conducted during warm summer days, and our laboratory was not equipped with air conditioning. Sweat is known to have an influence on ECG electrode conductance. It may also have an impact on PPG measurements by the examined wearable devices.

Data Annotation and Exclusion

Owing to various influencing factors—mainly ECG electrode loss, heat, selected activities, and other unknown skin factors—less data than anticipated were ultimately included in our study (20/23, 87% participants). A solid baseline (ground truth) was of the utmost importance in our study. Our manual data annotation of the criterion device data underlines this effort. As the annotation affects only the criterion device, it has no impact on the data recorded by the evaluated devices and, therefore, on future studies.

For participants 2, 5, and 7, only a subset of activities was included in our statistical analysis (Table 2). Although the respective individual activity metric averages reported in Table 4 do not include data for the respective activities, we did not exclude these individual participants for the overall metrics. This may lead to a minor bias toward resting and sedentary activities, as activities with higher physical activities were more prone to noise and, thus, data exclusion. Metrics only show minor changes if the data of these participants are excluded from the overall metric. The overall Fitbit Charge 4 MAE changed from 8.589 to 8.614 upon exclusion, and the Samsung Galaxy Watch Active2 MAE increased from 8.13 to 8.429.

The inclusion of data of participant 20 is controversial. A main argument for potential exclusion is that the data are clearly erroneous, and such data would be equally excluded in the study settings. On the other hand, faulty recordings may also occur in real-life settings. Excluding the data would lead to a positive bias in favor of the Samsung Galaxy Watch Active2 and, thus, to a nonobjective comparison. Therefore, we decided to include these data.

Conclusions

We evaluated 2 previously unvalidated wearable devices by conducting a study featuring various activities and 23 participants. Throughout the entire experimental procedure, both devices achieved results just <10% MAPE and thus, presented acceptable HR measurement capabilities. The Fitbit Charge 4 outperformed the Samsung Galaxy Watch Active2 during resting and sedentary activities, and the Samsung device was more accurate during high-intensity activities. Neither device reached sufficient accuracy during seated rest and keyboard typing.

Our study was a prequel to a larger interdisciplinary study in obstetrics. Researchers should consider the intended use of wearable devices when reviewing validation studies and evaluating their respective findings with respect to their full requirements. This is not only the case for the experimental design but also for other aspects. Accuracy may not be the only decisive factor. Features such as raw data access, battery runtime, or additional sensors may be equally relevant for individual research.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Zheng YL, Ding XR, Poon CC, Lo BP, Zhang H, Zhou XL, et al. Unobtrusive sensing and wearable devices for health informatics. IEEE Trans Biomed Eng 2014;61(5):1538-1554 [FREE Full text] [doi: 10.1109/TBME.2014.2309951] [Medline: 24759283]

2. Nelson BW, Low CA, Jacobson N, Areán P, Torous J, Allen NB. Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. NPJ Digit Med 2020;3:90 [FREE Full text] [doi: 10.1038/s41746-020-0297-4] [Medline: 32613085]

3. Daskivich TJ, Houman J, Lopez M, Luu M, Fleshner P, Zaghiyan K, et al. Association of wearable activity monitors with assessment of daily ambulation and length of stay among patients undergoing major surgery. JAMA Netw Open 2019;2(2):e187673 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.7673] [Medline: 30707226]

4. Lee JE, Lee DH, Oh TJ, Kim KM, Choi SH, Lim S, et al. Clinical feasibility of monitoring resting heart rate using a wearable activity tracker in patients with thyrotoxicosis: prospective longitudinal observational study. JMIR Mhealth Uhealth 2018;6(7):e159 [FREE Full text] [doi: 10.2196/mhealth.9884] [Medline: 30006328]

5. Goodale BM, Shilaih M, Falco L, Dammeier F, Hamvas G, Leeners B. Wearable sensors reveal menses-driven changes in physiology and enable prediction of the fertile window: observational study. J Med Internet Res 2019;21(4):e13404 [FREE Full text] [doi: 10.2196/13404] [Medline: 30998226]

6. Shilaih M, Goodale BM, Falco L, Kübler F, De Clerck V, Leeners B. Modern fertility awareness methods: wrist wearables capture the changes in temperature associated with the menstrual cycle. Biosci Rep 2018;38(6):BSR20171279 [FREE Full text] [doi: 10.1042/BSR20171279] [Medline: 29175999]

7. Shilaih M, de Clerck V, Falco L, Kübler F, Leeners B. Pulse rate measurement during sleep using wearable sensors, and its correlation with the menstrual cycle phases, a prospective observational study. Sci Rep 2017;7(1):1294 [FREE Full text] [doi: 10.1038/s41598-017-01433-9] [Medline: 28465583]

8. Mishra T, Wang M, Metwally AA, Bogu GK, Brooks AW, Bahmani A, et al. Pre-symptomatic detection of COVID-19 from smartwatch data. Nat Biomed Eng 2020;4(12):1208-1220. [doi: 10.1038/s41551-020-00640-6] [Medline: 33208926]

9. Quer G, Radin JM, Gadaleta M, Baca-Motes K, Ariniello L, Ramos E, et al. Wearable sensor data and self-reported symptoms for COVID-19 detection. Nat Med 2021;27(1):73-77. [doi: 10.1038/s41591-020-1123-x] [Medline: 33122860]

10. Schembre SM, Liao Y, Robertson MC, Dunton GF, Kerr J, Haffey ME, et al. Just-in-time feedback in diet and physical activity interventions: systematic review and practical design framework. J Med Internet Res 2018;20(3):e106 [FREE Full text] [doi: 10.2196/jmir.8701] [Medline: 29567638]

11. Fawcett E, Van Velthoven MH, Meinert E. Long-term weight management using wearable technology in overweight and obese adults: systematic review. JMIR Mhealth Uhealth 2020;8(3):e13461 [FREE Full text] [doi: 10.2196/13461] [Medline: 32154788]

12. Sartor F, Papini G, Cox LG, Cleland J. Methodological shortcomings of wrist-worn heart rate monitors validations. J Med Internet Res 2018;20(7):e10108 [FREE Full text] [doi: 10.2196/10108] [Medline: 29967000]

13. Kristoffersson A, Lindén M. Wearable sensors for monitoring and preventing noncommunicable diseases: a systematic review. Information 2020;11(11):521. [doi: 10.3390/info11110521]

14. Van Hooren B, Goudsmit J, Restrepo J, Vos S. Real-time feedback by wearables in running: current approaches, challenges and suggestions for improvements. J Sports Sci 2020;38(2):214-230. [doi: 10.1080/02640414.2019.1690960] [Medline: 31795815]

15. Kroll RR, McKenzie ED, Boyd JG, Sheth P, Howes D, Wood M, WEARable Information Technology for hospital INpatients (WEARIT-IN) study group. Use of wearable devices for post-discharge monitoring of ICU patients: a feasibility study. J Intensive Care 2017;5:64 [FREE Full text] [doi: 10.1186/s40560-017-0261-9] [Medline: 29201377]

16. Lee JM, An HS, Kang SK, Kim Y, Dinkel D. Examining the validity of Fitbit charge HR for measuring heart rate in free-living conditions. Med Sci Sports Exerc 2016;48(5S):786-787. [doi: 10.1249/01.mss.0000487361.48518.aa]

17. Brazendale K, Decker L, Hunt ET, Perry MW, Brazendale AB, Weaver RG, et al. Validity and wearability of consumer-based fitness trackers in free-living children. Int J Exerc Sci 2019;12(5):471-482 [FREE Full text] [Medline: 30899354]

18. Jo E, Lewis K, Directo D, Kim MJ, Dolezal BA. Validation of biofeedback wearables for photoplethysmographic heart rate tracking. J Sports Sci Med 2016;15(3):540-547 [FREE Full text] [Medline: 27803634]

19. Reddy RK, Pooni R, Zaharieva DP, Senf B, El Youssef J, Dassau E, et al. Accuracy of wrist-worn activity monitors during common daily physical activities and types of structured exercise: evaluation study. JMIR Mhealth Uhealth 2018;6(12):e10338 [FREE Full text] [doi: 10.2196/10338] [Medline: 30530451]

20. Thomson EA, Nuss K, Comstock A, Reinwald S, Blake S, Pimentel RE, et al. Heart rate measures from the Apple Watch, Fitbit Charge HR 2, and electrocardiogram across different exercise intensities. J Sports Sci 2019;37(12):1411-1419. [doi: 10.1080/02640414.2018.1560644] [Medline: 30657025]

21. Benedetto S, Caldato C, Bazzan E, Greenwood DC, Pensabene V, Actis P. Assessment of the Fitbit Charge 2 for monitoring heart rate. PLoS One 2018;13(2):e0192691 [FREE Full text] [doi: 10.1371/journal.pone.0192691] [Medline: 29489850]

22. Muggeridge DJ, Hickson K, Davies AV, Giggins OM, Megson IL, Gorely T, et al. Measurement of heart rate using the polar OH1 and Fitbit charge 3 wearable devices in healthy adults during light, moderate, vigorous, and sprint-based exercise: validation study. JMIR Mhealth Uhealth 2021;9(3):e25313 [FREE Full text] [doi: 10.2196/25313] [Medline: 33764310]

23. Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of heart rate watches: implications for weight management. PLoS One 2016;11(5):e0154420 [FREE Full text] [doi: 10.1371/journal.pone.0154420] [Medline: 27232714]

24. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017;7(2):3 [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]

25. El-Amrawy F, Nounou MI. Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? Healthc Inform Res 2015;21(4):315-320 [FREE Full text] [doi: 10.4258/hir.2015.21.4.315] [Medline: 26618039]

26. Düking P, Fuss FK, Holmberg HC, Sperlich B. Recommendations for assessment of the reliability, sensitivity, and validity of data provided by wearable sensors designed for monitoring physical activity. JMIR Mhealth Uhealth 2018;6(4):e102 [FREE Full text] [doi: 10.2196/mhealth.9341] [Medline: 29712629]

27. Parak J, Korhonen I. Evaluation of wearable consumer heart rate monitors based on photopletysmography. Annu Int Conf IEEE Eng Med Biol Soc 2014;2014:3670-3673. [doi: 10.1109/EMBC.2014.6944419] [Medline: 25570787]

28. Navalta JW, Montes J, Bodell NG, Salatto RW, Manning JW, DeBeliso M. Concurrent heart rate validity of wearable technology devices during trail running. PLoS One 2020;15(8):e0238569 [FREE Full text] [doi: 10.1371/journal.pone.0238569] [Medline: 32866216]

29. Thomas S, Reading J, Shephard RJ. Revision of the physical activity readiness questionnaire (PAR-Q). Can J Sport Sci 1992;17(4):338-345. [Medline: 1330274]

30. Fitbit charge 4: advanced fitness tracker. Fitbit. URL: https://www.fitbit.com/global/us/products/trackers/charge4 [accessed 2021-01-08]

31. Samsung India. URL: https://www.samsung.com/in/microsite/galaxy-watch-active2/ [accessed 2021-01-08]

32. Moon JH, Kang MK, Choi CE, Min J, Lee HY, Lim S. Validation of a wearable cuff-less wristwatch-type blood pressure monitoring device. Sci Rep 2020;10(1):19015 [FREE Full text] [doi: 10.1038/s41598-020-75892-y] [Medline: 33149118]

33. Galaxy watch active 2. Samsung de. URL: https://www.samsung.com/de/wearables/galaxy-watch-active2/ [accessed 2021-01-08]

34. Get started with Tizen wearable applications. Tizen docs. URL: https://docs.tizen.org/application/profiles/wearable/ [accessed 2021-01-08]

35. Biofeedback, neurofeedback and qEEG. Mind Media - BioTrace+ Software. URL: https://www.mindmedia.com/en/products/biotrace-software/ [accessed 2021-01-08]

36. Welcome to BioSPPy - BioSPPy 0.6.1 documentation. BioSPPy. URL: https://biosppy.readthedocs.io/en/stable/ [accessed 2021-01-08]

37. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature 2020;585(7825):357-362 [FREE Full text] [doi: 10.1038/s41586-020-2649-2] [Medline: 32939066]

38. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020;17(3):261-272 [FREE Full text] [doi: 10.1038/s41592-019-0686-2] [Medline: 32015543]

39. McKinney W. Data structures for statistical computing in Python. In: 9th Python in Science Conference. 2010 Presented at: SciPy '10; June 28-30, 2010; Austin p. 56-61. [doi: 10.25080/majora-92bf1922-00a]

40. Vallat R. Pingouin: statistics in Python. J Open Source Softw 2018;3(31):1026. [doi: 10.21105/joss.01026]

41. GitHub. stylianos-kampakis/supervisedPCA-Python. URL: https://github.com/stylianos-kampakis/supervisedPCA-Python [accessed 2021-04-22]

42. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989;45(1):255-268. [doi: 10.2307/2532051]

43. Nelson BW, Allen NB. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: intraindividual validation study. JMIR Mhealth Uhealth 2019;7(3):e10828 [FREE Full text] [doi: 10.2196/10828] [Medline: 30855232]

44. Baek S, Ha Y, Park HW. Accuracy of wearable devices for measuring heart rate during conventional and Nordic walking. PM R 2021;13(4):379-386. [doi: 10.1002/pmrj.12424] [Medline: 32449291]

XSL·FO
RenderX

45.   Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med 2018;18(3):91-93 [FREE Full text] [doi: 10.1016/j.tjem.2018.08.001] [Medline: 30191186]

Abbreviations

API: application programming interface
bpm: beats per minute
CCC: Lin concordance correlation coefficient
ECG: electrocardiogram
HR: heart rate
MAE: mean absolute error
MAPE: mean absolute percentage error
PCC: Pearson correlation coefficient
PPG: photoplethysmography or photoplethysmogram

XSL•FO
RenderX

Original Paper

Barriers to and Facilitators of Using a One Button Tracker and Web-Based Data Analytics Tool for Personal Science: Exploratory Study

Tom H van de Belt[1], PhD; Aimee de Croon[1], BSc; Faye Freriks[1], MSc; Thomas Blomseth Christiansen[2], MSc; Jakob Eg Larsen[3], PhD; Martijn de Groot[1], PhD

[1]Health Innovation Labs, Radboud University Medical Center, Nijmegen, Netherlands
[2]Konsulent Blomseth, Copenhagen, Denmark
[3]Danmarks Tekniske Universitet Compute (DTU), Technical University of Denmark, Copenhagen, Denmark

Corresponding Author:
Martijn de Groot, PhD
Health Innovation Labs
Radboud University Medical Center
Geert Grooteplein Noord 15
Nijmegen, 6525 EZ
Netherlands
Phone: 31 613424584
Email: martijn.degroot@radboudumc.nl

Abstract

Background: Individuals' self-tracking of subjectively experienced phenomena related to health can be challenging, as current options for instrumentation often involve too much effort in the moment or rely on retrospective self-reporting, which is likely to impair accuracy and compliance.

Objective: This study aims to assess the usability and perceived usefulness of low-effort, in-the-moment self-tracking using simple instrumentation and to establish the amount of support needed when using this approach.

Methods: In this exploratory study, the One Button Tracker—a press-button device that records time stamps and durations of button presses—was used for self-tracking. A total of 13 employees of an academic medical center chose a personal research question and used the One Button Tracker to actively track specific subjectively experienced phenomena for 2 to 4 weeks. To assess usability and usefulness, we combined qualitative data from semistructured interviews with quantitative results from the System Usability Scale.

Results: In total, 29 barriers and 15 facilitators for using the One Button Tracker were found. Ease of use was the most frequently mentioned facilitator. The One Button Tracker's usability received a median System Usability Scale score of 75.0 (IQR 42.50), which is considered as good usability. Participants experienced effects such as an increased awareness of the tracked phenomenon, a confirmation of personal knowledge, a gain of insight, and behavior change. Support and guidance during all stages of the self-tracking process were judged as valuable.

Conclusions: The low-effort, in-the-moment self-tracking of subjectively experienced phenomena has been shown to support personal knowledge gain and health behavior change for people with an interest in health promotion. After addressing barriers and formally validating the collected data, self-tracking devices may well be helpful for additional user types or health questions.

KEYWORDS

XSL•FO
RenderX

Introduction

Personal Science

An increasing number of people collect data on personal health or lifestyle phenomena, such as physical activity levels, mood, or sleep quality, for the purpose of self-reflection or personal knowledge gain. This practice of *self-tracking* or *self-quantification* can be supported with technological instruments such as wearable devices or mobile apps [1]. Enhancing personal knowledge supports health maintenance and may facilitate behavioral change [2-4]. In a clinical context, self-tracked data may be used for shared decision-making because it can improve communication, enhance bilateral coordination of care, and boost patient engagement and sense of autonomy [5-7]. Effective self-tracking not only mediates personal health behavior but also holds promise as a complementary field of knowledge creation and discovery. As a data acquisition method, self-tracking is integral to participant-led research (PLR) and personal science [8-11]. In PLR and personal science, patients or participants transcend their traditional role as a source of data by initiating and conducting research projects themselves. This includes acquisition and reflection of personally relevant data by and for themselves. Such an active role of a patient or citizen leading their own research is what defines and also connects PLR and personal science.

Self-tracking

In self-tracking, objective measures of an increasing range of physiological and behavioral phenomena can be automatically and accurately recorded by wearable sensors incorporated in activity trackers, smart watches, or other self-tracking devices. However, tracking subjectively experienced phenomena that manifest as physical sensations such as different moods, stress, discomfort, pain, mental flow, thoughts, emotions, or social interaction remains challenging. Instruments aimed at tracking these subjective phenomena as a primary or secondary outcome measurement using empirical methods typically involve diary entry or experience sampling methods with prompted self-reports, which can lead to inaccurate data due to associated memory recall biases [12-14]. Alternative options for in-the-moment registration, such as paper-based tracking, involve high effort, which may impair sustained use. These barriers limit discovery using self-tracking of subjectively experienced phenomena. To overcome these barriers, self-tracking devices that facilitate low-effort, in-the-moment tracking have attracted research interest. A pilot study exploring the use of a smart button device for the purpose of self-tracking medication adherence found that participants generally considered the device acceptable to use, but the collected data had poor concordance with electronic data collection [15].

One Button Tracker and Objectives

In recent case studies, the use of One Button Tracker instrumentation was introduced to facilitate low-effort, in-the-moment self-tracking [16,17]. The One Button Tracker is a data acquisition instrument that allows users to track any subjectively experienced phenomenon in the moment it occurs with little effort by a push of the single button. The point in time and duration of the button press are recorded, and the acquired data points can be loaded into a web-based data analytics tool. There, the collected data are automatically displayed in a calendar overview and graphs showing hourly and weekly distribution of observations. In one study [16], a patient with posttraumatic stress disorder was able to learn about the nature of his symptoms and the conditions in which they would arise by tracking a subjectively experienced precursor to one of his symptoms. In another study [17], the One Button Tracker enabled the investigation of temporal dynamics of and relations between two different subjectively experienced symptoms (intrusions and ruminations related to posttraumatic stress disorder). These studies suggest that the One Button Tracker in combination with a web-based data analytics tool can support low-effort, in-the-moment self-tracking. However, this promise remains a matter of research. Therefore, this exploratory study has two aims: to assess the usability and perceived usefulness of the One Button Tracker as instrumentation and to establish the amount of support needed using this approach.

Methods

Study Design and Setting

In this pilot study, we used a mixed methods approach to assess the usability and perceived usefulness of low-effort, in-the-moment tracking using simple instrumentation and to evaluate the support needed to allow individuals to self-track effectively. The qualitative part of the study consisted of semistructured interviews exploring facilitators and barriers to the use of this self-tracking method, perceived positive and negative effects of its use, and participant views on the potential value of support during the self-tracking process. The quantitative data were generated through a usability survey. Data were collected from February to August 2020. Ethical approval was granted by the local medical ethics committee (Medisch Ethische Toetsingscommissie Oost-Nederland, review number 2019-6066) and standards for reporting qualitative research were followed [18]. All participants provided written informed consent before participating in the study.

Participants

Participants were recruited from an existing cohort of employees from Radboud University Medical Center (Radboudumc, Nijmegen, the Netherlands) enrolled in the hospital's health promotion program *Healthy Professionals*. This program strives to educate employees on health and lifestyle topics to help them remain resilient and healthy in the rapidly changing health care environment [19]. This study was embedded within the Healthy Professionals program as self-tracking may aid participants in working toward their formulated lifestyle goals and because of the pioneering character of the program, which suited the explorative character of this study well. All 101 health care professionals enrolled in the program at the time of recruitment (May 2020) were invited to participate in the study by email. If they expressed interest, a researcher (AC) contacted them by phone to determine eligibility and answer any questions regarding the study's protocol. Applicants were eligible if they were aged ≥18 years. They were excluded if they were not able

to verbally communicate in Dutch or had cognitive dysfunction. We aimed to include 12-15 participants, which was deemed as sufficient for this exploratory study.

The One Button Tracker

For this study, we used a research prototype of the One Button Tracker, as depicted in Figure 1. The instrument was invented and developed by the coauthors JEL and TBC. The instrument is 41×31×12.5 mm in size, consists of a low-powered (3.3 V) press-button tracker within a 3D printed plastic casing, and can be charged via a USB port. When pressed, the device vibrates, thereby providing haptic feedback to the user. The point in time and duration of the button press were recorded. In the processing of the acquired data, these attributes can be used to distinguish between single, double, or more presses and shorter and longer durations of the presses, which can be used for different purposes depending on the user's needs. The One Button Tracker has no wireless or internet connection to ensure privacy, and the user is in control of the acquired data. The time required for a user to record an observation using the instrument was <1 second.

Figure 1. One Button Tracker. It is a data acquisition instrument in a 3D printed plastic casing that can be charged via a USB port. It was designed to track any subjectively experienced phenomenon.

Data stored on the device can be accessed by connecting the One Button Tracker to a laptop or desktop computer via a USB connection. The stored data file containing timestamps and button press durations can be loaded into a web-based data analytics tool. There, the collected data are automatically displayed in an overview table. In addition, several graphs are created that portray the average number of button presses per hour, day, week, and month (examples of visualizations from the data analytics tool are shown in Multimedia Appendix 1).

Study Procedures

Participants were sent instructions on the use of the One Button Tracker and web-based data analytics tools by email. These detailed how to operate the device; explained its ability to distinguish between 1, 2, and 3 presses; and instructed participants to charge the device twice a week. In addition, the hyperlink to the web-based data analytics tool was provided, accompanied by written instructions in a step-by-step manner on how to transfer data to the tool. In an intake session, these instructions were restated, and remaining questions were answered.

Thereafter, participants worked toward a suitable personal research question with the help of a researcher (AC). If participants had already formulated questions related to their health promotion goals before the Healthy Professionals intake session, these questions were examined to determine their suitability. If formulating a question was revealed to be difficult, the researcher used open-ended questions to explore the knowledge gaps in the path toward their formulated health goals.

A research question was deemed suitable if it related to the participants' set health or lifestyle goals, was personally relevant, and was answerable by self-tracking a given phenomenon. Participants were strongly advised to pick a phenomenon that would result in ≥2 and ≤20 clicks per day to avoid tracking fatigue.

Subsequently, participants started the self-tracking process in which they used the One Button Tracker to actively track the chosen phenomenon for a minimum of 2 and a maximum of 4 weeks, depending on their research question and preferences. During this time, participants could view the collected data through the data analytics tool on their computer when desired. In case of problems, they could contact a technical support helpline by phone or email. In addition, we performed a weekly checkup by phone to answer any questions and gauge the use of the device. Any remarks regarding the One Button Tracker or the self-tracking process participants made in these calls were noted and added to the record for analysis.

When participants finished tracking, they handed the device back to the researchers. In individual interviews, their experiences using the One Button Tracker and doing self-tracking were explored. In these sessions, the data collected by the participants were loaded into the web-based data analytics tool and discussed. Afterward, all data were deleted permanently. Following the interview, participants were asked to fill out the System Usability Scale (SUS) questionnaire to determine the usability of the One Button Tracker. Figure 2 provides an overview of the study's timeline.

Figure 2. Overview of the study's timeline. Self-tracking was carried out by the participants for a duration of 2 to 4 weeks.
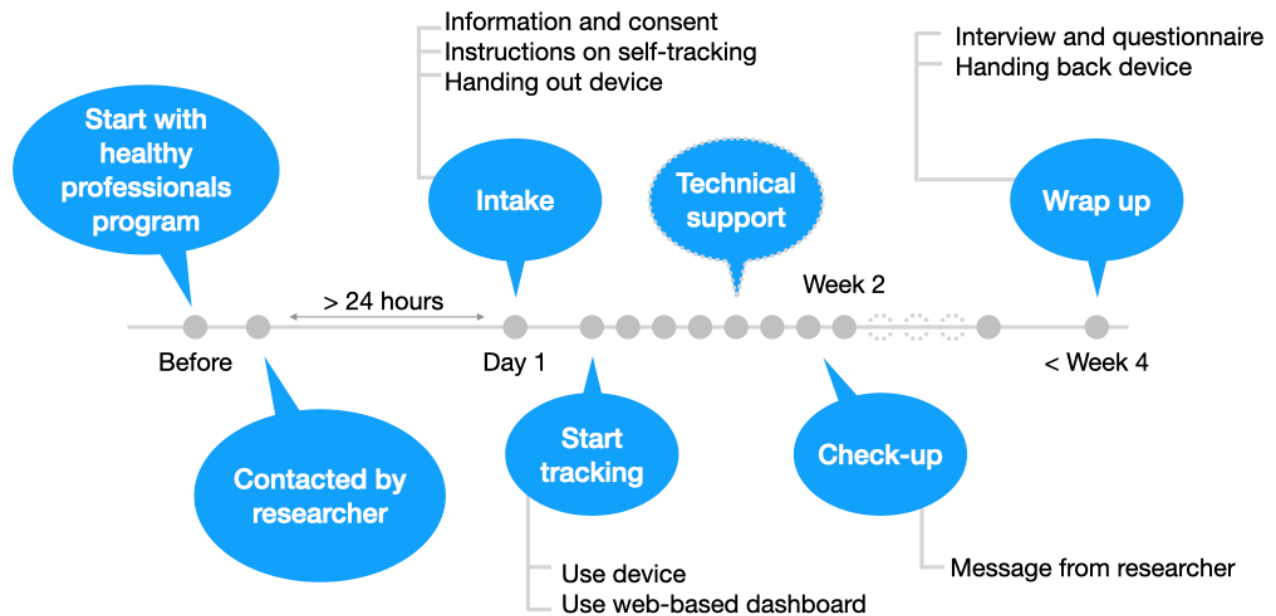
Data Collection

All semistructured interviews were conducted in July 2020 by a junior researcher (AC) with previous training in qualitative research. The interview guide was developed specifically for this study, drawing from insights from Li et al [20] and Almalki et al [21] regarding personal informatics systems. It included questions on participant expectations, their personal research question and approach, views of and experiences with the One Button Tracker and the web-based data analytics tool, and views of and experiences with self-tracking and support during this process (Multimedia Appendix 2). All interview questions were open-ended. In addition to these questions, participants were asked to grade the OBT on a scale from 1 (worst possible functioning) to 10 (best possible functioning).

During the interview sessions, we also collected demographic information, including birth year, education, and job description. We initially intended to interview all participants face-to-face at the participants' location of choice. However, owing to COVID-19 restrictions, half of the interviews were conducted via the internet using Skype for Business (version 7.0.2676.1, 2018; Microsoft Corporation). All interviews were audio-recorded, transcribed clean verbatim, and anonymized.

Following the interview, participants completed a Dutch version of the SUS [22] (Multimedia Appendix 3). This short 10-item validated questionnaire is widely used to determine the usability of devices or systems. It asks participants to rate the truthiness of 10 statements concerning the device's usability on a 5-point Likert scale.

Data Analysis

A total of 2 researchers (AC and FF) analyzed the anonymized transcripts independently using qualitative data analysis software (ATLAS.ti version 7.1; Scientific Software Development GmbH). They identified barriers, facilitators, and positive and negative effects of the use of the One Button Tracker for self-tracking. The identified codes were thoroughly discussed

until consensus was reached. Remaining disagreements were discussed with a third researcher (TB). To determine if the sample size was sufficient to gain a comprehensive overview of participant experiences, code saturation was assessed after each third interview by examining whether any previously unnamed barriers and facilitators or effects were identified in the newly gathered data. We defined saturation as 3 subsequent interviews with no new factors.

Facilitators and barriers regarding the use of the One Button Tracker were categorized according to the Gagnon framework concerning determinants of adoption of information and communications technologies in health [23,24]. New barriers and facilitators were added to the framework. The Donabedian framework for the quality of health care was used to present all identified positive and negative effects [25]. This framework distinguishes structure (context in which health care is delivered), process (all actions that make up health care), and outcome (all effects on patients' health). Dutch quotes and themes used in this paper were translated into English and checked by all authors.

Participants' answers on the SUS were computed as described by Brooke et al [22], resulting in a score between 0 and 100. Scores were interpreted in accordance to Bangor et al [26], where with a score of ≤50.9, the usability of a device is deemed *poor*; with a score of >50.9, usability is deemed *sufficient*; with a score of >71.4, usability is deemed *good*; with a score of >85.5, usability is deemed *excellent*; and with a score of >90.9, usability is deemed the *best imaginable*.

Analysis and statistics were performed in the Radboudumc using SPSS (version 25.0; IBM). Normally distributed continuous variables are described using mean and SD. Median and interquartile values were shown in case variables were not normally distributed. Qualitative or categorical variables were described using frequencies and percentages.

Data Storage and Privacy

After each interview, the pseudonymized audio record was stored within the Radboudumc data storage environment until transcription was finished, in accordance with Dutch privacy law. After transcription and coding, the anonymized transcripts were archived according to the Radboudumc research policy.

Results

Participants

Of the 101 professionals who were invited to participate, 58.4% (59/101) actively declined. Provided reasons for not participating included no time, increased workload due to the COVID-19 pandemic (7/59, 12%) already able to reach health goals (3/59, 5%), not wanting to use this particular self-quantification system (3/59, 5%), changing jobs (2/59, 2%), and illness (2/59, 2%). A total of 13 professionals expressed interest and were included, resulting in a recruitment rate of 12.8% (13/101). Table 1 provides an overview of all participants' characteristics.

Table 1. Participant characteristics (n=13).

Characteristic	Value
Gender, n (%)	
Female	11 (85)
Male	2 (15)
Age (years), median (range)	56 (35-67)
Educational background, n (%)	
Vocational	3 (23)
Applied sciences	6 (46)
Academic	4 (31)
Job type, n (%)	
Management	7 (54)
Medical staff	
Nursing and care	3 (23)
Medical doctor and specialists	0 (0)
Staff, administration, or secretary	3 (23)

Self-tracking

All participants successfully formulated a personal research question and used the One Button Tracker for self-tracking. A total of 85% (11/13) of the participants made use of the device's ability to distinguish between 1, 2, and 3 button presses. Some used this feature to indicate the intensity of the tracked phenomenon, where more presses indicated a stronger or more intense experience, whereas others assigned different relevant phenomena to different numbers of presses to explore their relationships. Table 2 provides an overview of all personal research questions and tracked phenomena. When participants were asked to restate their formulated personal question during the interviews, they often posed their focus as a behavioral goal or an aim, rather than as a (research) question.

Table 2. Topic, question, and description of the self-tracked phenomenon of each of the 13 participants' personal research projects.

ID	Topic	Personal research question	Phenomenon description
01	Creativity	How can I identify and facilitate creative moments in my day-to-day work?	1 press=creative moment; 2 or 3 presses=more intense experience
02	Snacking	How often do I snack?	1 press=snack
03	Stress	What factors influence my (experience of) heart palpitations?	1 press=palpitations; 2 or 3 presses=more intense experience (used diary to record contextual factors)
04	Stress	To what extent do I experience physical stress symptoms?	1 press=muscle tension; 2 presses=tensed breathing; 3 presses=cannot eat
05	Stress	What is for me the relation between stress and movement?	1 press=stress; 2 presses=moving a little; 3 presses=moving a lot
06	Hydration	How much do I drink and how does this relate to thirst?	1 press=thirst; 2 presses=drinking
07	Creativity	How can I get out of a rut, using music?	1 press=effective music-based intervention
08	Hydration	How much do I drink and how does this influence my headaches?	1 press=drinking; 2 presses=headache; 3 presses=pain killers
09	Drinking	How much coffee and wine do I consume?	1 press=cup of coffee; 2 presses=glass of wine
10	Wellness	At what times do I feel energetic?	More presses=feeling better; fewer or no presses=feeling worse
11	Disquiet and food	What is for me the relationship between mental unrest and thinking of food?	1 press=thinking of food; 2 presses=mental unrest
12	Fidgeting	Does my fidgeting habit follow a recognizable pattern, and how might knowledge of this pattern help me to fidget less?	1 press=fidgeting with fingers; 2 presses=touching face
13	Hydration and stillness	How much do I drink, and how can I intercept my working day with mindfulness exercises to improve stillness?	1 press=drinking; 2 presses=mindfulness exercise

The duration of the self-tracking projects ranged from 6 to 38 days, with a median of 23. Of the 13 participants, 3 (23%) deviated from the self-tracking period of 2 to 4 weeks. From these 13 participants, 1 (8%) ended the tracking project after 8 days because of a perceived lack of usefulness and high burden, another (1/13, 8%) ended the project after 6 days because she had answered her personal research question, and the last (1/13, 8%) extended the tracking period because of illness. During the tracking period, 54% (7/13) of the participants visualized the collected data using the web-based data analytics tool. Moreover, of the 13 participants, 2 (15%) used the tool for data interpretation, whereas the other 5 (38%) used it to check whether the One Button Tracker was still fully operational. Of the 6 participants who did not successfully load their data into the web-based tool, 1 (17%) intended to but encountered technical issues, whereas the others had no interest in viewing the collected data.

Technical Challenges

Of the 13 participants, 6 (46%) participants encountered technical difficulties during the tracking period. After troubleshooting, it turned out that a technical error had resulted in the loss of all software and data files. In 3 cases, no personal data were lost, as the technical error had occurred before self-tracking commencement. However, 23% (3/13) of the participants lost (a proportion of) their collected data when this transpired.

Multiple possible causes for this issue were identified. Of the 13 participants, 1 (8%) had accidently connected the One Button Tracker to the hospital's computers, which had been warned against in the written instructions, as this was known to elicit a faulty reset of the One Button Tracker. Another (1/13, 8%) participant inadvertently caused an error by incorrectly detaching the One Button Tracker from their PC. In other cases, the cause of the errors remained unclear. After consultation with the coauthors JEL and TBC, all One Button Trackers received a software update. From there on, no further technical issues with the One Button Trackers were encountered.

Qualitative Results

The interviews ranged from 15 to 56 minutes in length. Analysis of the interviews revealed 29 barriers and 15 facilitators, as well as 2 negatively and 12 positively perceived effects. Textbox 1 presents all identified barriers and facilitators. Textbox 2 presents an overview of the experienced effects. Data saturation could not be confirmed.

Textbox 1. Barriers to and facilitators regarding the use of the One Button Tracker as a self-tracking instrument according to the Gagnon framework of determinants of adoption of information and communications technologies in health care. Barriers and facilitators are ordered in three categories (technological, individual, and external), including the number of participants who mentioned each identified barrier (B, n) or facilitator (F, n).

Technological barriers and facilitators related to mobile health characteristics

1. Design and technical concerns
   - Unattractiveness of design (B, 1)
   - Physical aspect of the device reminds user to track (F, 1)
   - Data analytics tool's visualizations are attractive (F, 1)
   - Device is too small to be aware of (B, 1)
   - Device's button (cannot be pressed or [lack of] confirmation; B, 1)
   - Device's button (vibrations assure user of correct press; F, 1)
   - Device is not compatible with operating system (B, 1)
   - Disfunction not further specified (B, 1)

2. Perceived usefulness
   - Set goal is personally important to user (F, 1)
   - Lack of usefulness after research question has been answered (B, 1)
   - No need to use device or dashboard (B, 4)
   - Lack of confidence that research question will be answered (B, 1)

3. Perceived ease of use
   - Device is handy (F, 3)
   - Device is easy to use (F, 5)
   - Carrying device is a bother: already have to bring a smartphone (B, 1)
   - Challenging to transport the device, especially when the user has no pockets (B, 5)
   - Visualizing the data in the data analytics tool involves too much effort (B, 3)
   - User needs instructions on how to interpret the data analytics tool's visualizations (B, 2)
   - Fear to lose device (B, 1)
   - Psychological stress of using the device is zero, which facilitates use (F, 1)

4. Privacy and security concerns
   - Data cannot be removed accidently (F, 1)
   - Good privacy protection when compared with other tools such as app (F, 1)

5. Satisfaction about content available (completeness)
   - Lack of in-the-moment feedback (B, 1)
   - Hard to determine when to use the device specifically with subjective phenomena (B, 1)

6. Content appropriate for users (relevance)
   - Sense of failure each time a negatively judged phenomena is tracked (B, 1)

7. Accuracy
   - User forgets to bring the device when on the move (B, 7)
   - User forgets to track when device is out of sight (B, 1)
   - User forgets to track when distracted by other activities (B, 1)
   - Incorrect categorization of measure due to changes midactivity or midexperience (B, 1)
   - Button cannot be pressed accidently (F, 1)
   - User forgets device when clothes are regularly changed, for example, health care workers (B, 1)

- After a few weeks user forgets to track (B, 1)

Individual barriers and facilitators: knowledge, attitude, and sociodemographic characteristics

1. Time issues

   - User is too busy to view the data analytics tool (B, 2)

   - Interpreting the data analytics tool visualizations is too time-consuming (B, 1)

2. Outcome expectancy

   - While working night shifts, life differs so substantially that tracking during these shifts does not lead to generalizable knowledge (B, 1)

3. Agreement with mHealth (welcoming or resistant)

   - Tracking is fun (F, 1)

   - Not motivated to use data analytics tool (B, 1)

External barriers and facilitators: social and training environment

1. Social pressure (associated with peers)

   - Device can be used unnoticed by others (F, 1)

   - Lack of use when other people are around (B, 1)

   - Less comfortable to use device among others: fear of having to explain him or herself or to disturb group progress (B, 1)

2. Training

   - Support and guidance during the process of designing a personal research question (F, 1)

   - No behavioral goal formally set in collaboration with researcher (B, 1)

3. Communication and collaboration effort

   - Regular checkups support user motivation (F, 1)

4. External environment

   - Research context elicits motivation (F, 2)

Textbox 2. Positive and negative effects of the use of the One Button Tracker as a self-tracking instrument according to the Donabedian framework for quality of care. Effects are ordered in two categories (process and outcome) including the number of participants who mentioned each positive effect (P, n) or negative effect (N, n).

Process

1. Self-tracking process

   - It is enjoyable to consciously focus on a certain experience or behavior (P, 2)

   - Tracking causes confrontation with failure in the set goal (N, 1)

   - The device causes annoyance (N, 1)

2. Tracked personal data

   - Gathered data can be related to lived experiences (P, 3)

   - The ability to differ between the number of presses facilitates a differentiated overview of the user's progress (P, 1)

   - Data analytics tool visualizations increase insight into temporal fluctuations in the tracked phenomenon (P, 1)

Outcome

1. Self-awareness

   - Enhanced awareness of tracked experience or phenomenon (P, 8)

2. Personal knowledge

   - (Objective) confirmation of existing beliefs (P, 3)

   - Gain of personal insights (P, 4)

   - Reassurance that user can take control of own health (P, 1)

3. Action

   - User is incentivized to turn goal setting into action (P, 4)

   - Behavioral change (P,4)

   - The device functions as an incentive to perform the desired behavior (P, 3)

   - The device functions as a reinforcement to put the gained insights into practice (P, 3)

We identified a diverse number of barriers and facilitators that influence the uptake of the One Button Tracker. However, 3 aspects of the One Button Tracker's usability were emphasized by participants. The *user-friendliness of the device* was one such aspect. Because of its small size and the simplicity of its design, the One Button Tracker was generally considered easy to use. Second, most participants thought of the One Button Tracker as *easily portable*, as users can carry the device with them in a trouser or shirt pocket. A third mentioned asset of the One Button Tracker was that its *size facilitates pressing its button unseen by others*, which was considered to be of added value:

> *I could just quickly press it, use it unnoticed. So I er, I didn't need to grab some clumsy-looking apparatus, "what have you got there?". You can just nicely...just quietly press it.* [Participant 03]

Although collecting personal data with the One Button Tracker was regarded to be convenient, there were certain aspects of the data collection process that participants viewed as possibly troublesome. Although the One Button Tracker was considered to be easily portable, it could be *challenging for users to remember bringing the device* when they were on the move. This effect was said to be more pronounced when participants'

clothing lacked pockets to transport the One Button Tracker. Even when participants brought the One Button Tracker with them, missed tracking points still occurred. Participants explained that they *sometimes forgot to track experiences when distracted by other activities* or when they had placed the One Button Tracker out of their line of sight. Because of these occurrences, of the 13 participants, 5 (38%) reported instances of missed or misregistered data points. However, even with these missed data points, the majority still felt that the collected data accurately represented their experienced reality:

> *One time, I wasn't wearing trousers with pockets. And then I realized, oh shoot, forgot it. No trousers, no tracker.* [Participant 05]

Another factor of which the importance was emphasized is the availability of tailored support during the tracking process. Without guidance, it would be difficult to design a personally relevant research question and to select an appropriate phenomenon to track. Furthermore, some participants felt that the points of contact with the researchers during the weekly checkups facilitated a boost in motivation to persist in the tracking process. Appropriate support was also judged to be essential for correct interpretation of collected data presented in the web-based data analytics tool. Overall, 31% (4/13) of the

participants mentioned that they were interested in looking at their personal data but needed support interpreting the data visualizations. Some stated they thought connecting the device to their desktop and loading the collected data involved too much effort or was too time-consuming. All in all, support and guidance during all stages of the self-tracking process were judged as very important:

> *Well, I did see those graphs. And I did have a look at them, but I thought yeah, I'm going to need some explanation from you here, what you guys think of this.* [Participant 03]

Participants experienced a range of effects to be the result of the use of the One Button Tracker for self-tracking. Although most of these effects were judged positively, of the 13 participants, 1 (8%) described that she experienced negative emotions throughout the tracking process. This participant had chosen to track a behavioral phenomenon that she wanted to perform less often, which resulted in a sense of failure each time this behavior was registered. As for positive effects, 62% (8/13) of the participants stated that the use of the One Button Tracker led to enhanced awareness of the tracked phenomenon. This facilitated an objective confirmation of existing beliefs but could also lead to the gain of entirely new personal insights. For 31% (4/13) of the participants, enhanced awareness and gained personal knowledge culminated in or contributed to behavioral change. Another important effect mentioned was the functioning of the One Button Tracker as an incentive for desired behavior. Participants explained that the device reminded them or even motivated them to act more in line with their stated behavioral goal:

> *Seeing that thing laying there, thinking oh right, oh I can do that now! That does have a certain action effect, so to speak.* [Participant 06]

Quantitative Results

The One Button Tracker received a median grade of 7.5 (IQR 2.0) on a scale of 1 to 10, with individual grades ranging from 5 to 10. The SUS was completed by all participants. The median SUS score was 75.0 (IQR 17.50) out of a possible maximum score of 100, with individual scores ranging from 50.0 to 97.5. This corresponds to a percentile rank of 73% and indicates that the One Button Tracker's usability can be considered as *good usability* [26,27].

Discussion

Principal Findings

In this study, we explored the potential of low-effort, in-the-moment self-tracking of subjectively experienced phenomena to support self-knowledge gain by focusing on one such option, the One Button Tracker.

All participants in this study successfully designed a personal research question in the context of a health promotion program and tracked one or more chosen phenomena with the device. The findings suggest that instrumentation options such as these can aid individuals in the pursuit of personal knowledge gain. However, such an approach may not suit everyone. Participants

highlighted the user-friendliness of the One Button Tracker instrument; however, some barriers to swift data collection were also identified.

Participants generally considered the One Button Tracker as user-friendly, which was reflected by the median SUS score and grade the device received. However, technical issues encountered initially by some of the participants posed a risk to loss of the collected personal data, and with it, the instrument's usability. Although such difficulties can occur in instruments that have not been tested extensively [15], problems of this type will have to be resolved through design testing before such self-tracking options are rolled out on a bigger scale [28,29]. Additional usability challenges associated with the physical design of the One Button Tracker, for instance, how to carry it when wearing clothes without pockets, were also a limitation.

Despite these barriers the One Button Tracker provided enhanced awareness and personal knowledge gain. Most participants felt using the One Button Tracker helped raise their awareness of the tracked phenomenon. They indicated that this awareness led to confirmation or gain of personal insights. This benefit was reported even though some participants experienced limited accuracy of the collected data caused by missed or misregistered observations. This suggests that high accuracy of the collected data may not be necessary to effectuate the positive effects, as almost all participants also reported instances of missed or misregistered data points. Further research is needed to evaluate how the accuracy of collected data might influence the quality of gained self-knowledge, a question that has been raised before [30], and how this quality might in turn influence users' perceptions of health or health behavior.

Most participants experienced a change in health behavior, yet the measurability and sustainability of this change remained unclear. In line with the *self-improvement hypothesis of personal informatics*, participants believed that the increased awareness and gained personal knowledge resulting from the self-tracking process led them to change their behavior [31,32]. In addition, some participants felt the change in behavior was incentivized specifically by the sight of the instrument. Importantly, it is an experience of behavioral change that was assessed here; we did not measure actual change.

Participants emphasized the value of support during the different stages of the self-tracking process. The need for support is exemplified by the fact that most participants in present and previous studies struggled with interpreting the collected data on their own [20,21]. Furthermore, in line with previous research, regular interaction with the researchers was stated to be a motivating factor [4,32]. However, evidence that such support contributes to enhanced health effects as compared with self-tracking with no assistance is lacking [29,33,34]. Interview data indicate that the need for support was often related to help with the web-based data analytics tools.

Implications

The diversity in participant views emphasizes the importance of providing an appropriate self-tracking option to each individual. It turned out that the low-effort in-the-moment

approach is a good fit for some individuals, but not all. Previous research supports the notion that a one-size-fits-all approach does not exist; rather, different options should be developed for user groups with different needs [35-37]. Self-tracking can lead to negative effects such as feelings of failure and incompetence associated with individuals being reminded of the incongruence between their behavioral goal and the behavior actually performed [32,37-40]. Therefore, it is important to determine what does and does not work for different users or user groups.

The results found here have several implications for scientific research and medical practice. Academic researchers are becoming more aware of the contributions PLR can make to knowledge creation, such as generating hypotheses, enriching questionnaires, or answering questions through real-life data. However, before low-effort, in-the-moment tracking can be used as a scientifically sound data acquisition method, more clarity is needed on the quality of the collected data. In addition to the barriers described above, the tracked experience may also be distorted because it is being tracked, a phenomenon that is addressed as the *observer effect* [32,41]. Therefore, further research is needed into the validity and reliability of the collected data, and how these relate to those of other data collection methods. In a clinical context, it is clear that self-tracking of personally relevant phenomena may help some individuals gain self-insight and develop healthier behavior. However, for such methods to be successfully deployed in medical practice, it is important to gain a better understanding of the possible health effects, long-term effects, and potential differences in experiences between different user groups, as it has been shown previously that patients often differ in experiences from healthy volunteers [30,42]. This will help determine in which application areas self-tracking could be beneficial and how such a program could be best set up.

This study is among the first to explore this novel approach to self-tracking. An important strength of the study is that it provided participants with a chance to participate in a personalized form of PLR, in which they could set up and answer their own personally relevant research question. This provided a real-life setting, in which participants self-tracked for a purpose of their own choosing. In addition, the use of a mixed methods approach enhanced the strength of the findings, as the quantitative and qualitative findings converged well.

Limitations

This study has several limitations. Data saturation was not reached. As a result, we cannot ensure that the entire range of possible participant experiences was covered. The participation rate was sufficient for saturation; however, it was relatively low (n=13). In our experimental setup, participants were allowed to track different types of phenomena, which explains the wide range of facilitators and barriers that were identified. Although

this small group already provides valuable information in the context of an explorative study, a larger group may have revealed additional barriers and facilitators or may have provided a clearer image about the ones that stand out. Another limitation is related to the Hawthorne effect. Some participants found the research context itself to be motivating. This could mean that participants are less motivated to keep up a real-life self-tracking project. Finally, the participants included here were highly motivated, with an active participation in a health promotion program. This suggests that their experiences might not fully correspond with those of a population with less strong motivation or fewer support options.

Concluding Remarks and Future Work

This study explored the potential usability and usefulness of low-effort, in-the-moment self-tracking for acquiring new personal insights and supporting changes in health behavior. Although the prototype self-tracking instrument studied here has shown itself to be perceived as user-friendly and can be used to quantify subjective experienced phenomena effectively, experiences in the participant group varied widely. Although the study has demonstrated the utility of the instrument in individual cases, the potential efficacy of the instrument in general is inconclusive at this point.

Before the One Button Tracker instrument can be provided to patients or study participants on a larger scale, the technical challenges and specific usability issues identified in this study should be addressed. In particular, usability issues related to wearability would need to be addressed.

As the study participants appreciated the support they received during the study, it would be interesting to study which kind of and what level of support is conducive for the process in the different steps of self-tracking as part of PLR or personal science. Further research is needed to assess the exact benefits of support and to evaluate how this support would best be provided. Here, the role and usability of the data analytics tool is also a topic for further research.

The explorative nature of this study and the overall purpose for using the One Button Tracker to track anything related to personal health and well-being has led participants to track a wide range of different phenomena. This demonstrated that participants can use the instrument in everyday life settings and acquire real-world data on subjective experience. Future studies could focus on addressing the application of the instrument in specific health domains to identify where self-tracking of subjective experience could potentially benefit diagnostics, health monitoring, or behavior change. For example, future studies could be conducted to understand to what extent the observed experiences would translate to objectively measurable change and how sustainable this change in behavior could be [14,31,32,43].

Conflicts of Interest

Multimedia Appendix 1
Overview with screenshots from the dashboard of the web-based analytics tool.
[DOCX File , 132 KB - formative_v6i3e32704_app1.docx]

Multimedia Appendix 2
Interview guide (in Dutch).
[DOCX File , 16 KB - formative_v6i3e32704_app2.docx]

Multimedia Appendix 3
Questions in the System Usability Scale (in Dutch).
[DOCX File , 20 KB - formative_v6i3e32704_app3.docx]

References

1. Swan M. The quantified self: fundamental disruption in big data science and biological discovery. Big Data 2013;1(2):85-99. [doi: 10.1089/big.2012.0002] [Medline: 27442063]
2. Burke LE, Swigart V, Warziski Turk M, Derro N, Ewing LJ. Experiences of self-monitoring: successes and struggles during treatment for weight loss. Qual Health Res 2009;19(6):815-828 [FREE Full text] [doi: 10.1177/1049732309335395] [Medline: 19365099]
3. Kong A, Beresford SA, Alfano CM, Foster-Schubert KE, Neuhouser ML, Johnson DB, et al. Self-monitoring and eating-related behaviors are associated with 12-month weight loss in postmenopausal overweight-to-obese women. J Acad Nutr Diet 2012;112(9):1428-1435 [FREE Full text] [doi: 10.1016/j.jand.2012.05.014] [Medline: 22795495]
4. Harris T, Kerry SM, Victor CR, Ekelund U, Woodcock A, Iliffe S, et al. A primary care nurse-delivered walking intervention in older adults: PACE (pedometer accelerometer consultation evaluation)-lift cluster randomised controlled trial. PLoS Med 2015;12(2):e1001783 [FREE Full text] [doi: 10.1371/journal.pmed.1001783] [Medline: 25689364]
5. Woods SS, Evans NC, Frisbee KL. Integrating patient voices into health information for self-care and patient-clinician partnerships: Veterans Affairs design recommendations for patient-generated data applications. J Am Med Inform Assoc 2016;23(3):491-495. [doi: 10.1093/jamia/ocv199] [Medline: 26911810]
6. Petersen C. Patient-generated health data: a pathway to enhanced long-term cancer survivorship. J Am Med Inform Assoc 2016;23(3):456-461. [doi: 10.1093/jamia/ocv184] [Medline: 26714765]
7. Kooiman TJ, de Groot M, Hoogenberg K, Krijnen WP, van der Schans CP, Kooy A. Self-tracking of physical activity in people with type 2 diabetes: a randomized controlled trial. Comput Inform Nurs 2018;36(7):340-349. [doi: 10.1097/CIN.0000000000000443] [Medline: 29742550]
8. Kaye J, Curren L, Anderson N, Edwards K, Fullerton SM, Kanellopoulou N, et al. From patients to partners: participant-centric initiatives in biomedical research. Nat Rev Genet 2012;13(5):371-376 [FREE Full text] [doi: 10.1038/nrg3218] [Medline: 22473380]
9. Vayena E, Tasioulas J. The ethics of participant-led biomedical research. Nat Biotechnol 2013;31(9):786-787. [doi: 10.1038/nbt.2692] [Medline: 24022150]
10. Grant AD, Wolf GI, Nebeker C. Approaches to governance of participant-led research: a qualitative case study. BMJ Open 2019;9(4):e025633 [FREE Full text] [doi: 10.1136/bmjopen-2018-025633] [Medline: 30944134]
11. Wolf GI, De Groot M. A conceptual framework for personal science. Front Comput Sci 2020;2:21. [doi: 10.3389/fcomp.2020.00021]
12. Altschuler A, Picchi T, Nelson M, Rogers JD, Hart J, Sternfeld B. Physical activity questionnaire comprehension: lessons from cognitive interviews. Med Sci Sports Exerc 2009;41(2):336-343 [FREE Full text] [doi: 10.1249/MSS.0b013e318186b1b1] [Medline: 19127192]
13. Archer E, Pavela G, Lavie CJ. The inadmissibility of what we eat in America and NHANES dietary data in nutrition and obesity research and the scientific formulation of national dietary guidelines. Mayo Clin Proc 2015;90(7):911-926 [FREE Full text] [doi: 10.1016/j.mayocp.2015.04.009] [Medline: 26071068]
14. Hermsen S, Frost J, Renes RJ, Kerkhof P. Using feedback through digital technology to disrupt and change habitual behavior: a critical review of current literature. Comput Hum Behav 2016;57:61-74 [FREE Full text] [doi: 10.1016/j.chb.2015.12.023]
15. Bartlett Ellis RJ, Hill JH, Kerley KD, Sinha A, Ganci A, Russell CL. The feasibility of a using a smart button mobile health system to self-track medication adherence and deliver tailored short message service text message feedback. JMIR Form Res 2019;3(2):e13558 [FREE Full text] [doi: 10.2196/13558] [Medline: 31237568]

16. Larsen J, Eskelund K, Christiansen TB. Active self-tracking of subjective experience with a one-button wearable: a case study in military PTSD. In: Proceedings of the 2nd Computing and Mental Health workshop at the ACM CHI. 2017 Presented at: CHI '17; May 6-11, 2017; Denver URL: https://arxiv.org/pdf/1703.03437.pdf

17. Arendt IT, Riisager LH, Larsen JE, Christiansen TB, Moeller SB. Distinguishing between rumination and intrusive memories in PTSD using a wearable self-tracking instrument: a proof-of-concept case study. Cogn Behav Ther 2021;14:E15. [doi: 10.1017/s1754470x2100012x]

18. Tong AP, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]

19. Radboud University Medical Center, Nijmegen, the Netherlands. Lifeguard. URL: https://lifeguard.nl/klanten/radboudumc/ [accessed 2022-02-01]

20. Li I, Dey A, Forlizzi J. A stage-based model of personal informatics systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2010 Presented at: CHI '10; April 10 - 15, 2010; Atlanta p. 557-566. [doi: 10.1145/1753326.1753409]

21. Almalki M, Gray K, Martin-Sanchez FJ. Refining the concepts of self-quantification needed for health self-management. A thematic literature review. Methods Inf Med 2017;56(1):46-54. [doi: 10.3414/ME15-02-0007] [Medline: 27523820]

22. Brooke J. SUS: a 'Quick and Dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability evaluation in industry. London: Taylor and Francis; 1996:189-194.

23. Gagnon MP, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]

24. Gagnon MP, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]

25. Donabedian A. The quality of care. How can it be assessed? JAMA 1988;260(12):1743-1748. [doi: 10.1001/jama.260.12.1743] [Medline: 3045356]

26. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123 [FREE Full text]

27. Sauro J, Lew JR. Quantifying the user experience: practical statistics for user research. Burlington: Morgan Kaufmann; 2012.

28. Kim J. A qualitative analysis of user experiences with a self-tracker for activity, sleep, and diet. Interact J Med Res 2014;3(1):e8 [FREE Full text] [doi: 10.2196/ijmr.2878] [Medline: 24594898]

29. Lentferink AJ, Oldenhuis HK, de Groot M, Polstra L, Velthuijsen H, van Gemert-Pijnen JE. Key components in eHealth interventions combining self-tracking and persuasive eCoaching to promote a healthier lifestyle: a scoping review. J Med Internet Res 2017;19(8):e277 [FREE Full text] [doi: 10.2196/jmir.7288] [Medline: 28765103]

30. Crawford K, Lingel J, Karppi T. Our metrics, ourselves: a hundred years of self-tracking from the weight scale to the wrist wearable device. Eur J Cult Stud 2015;18(4-5):479-496. [doi: 10.1177/1367549415584857]

31. Kersten-van Dijk ET, Westerink JH, Beute F, IJsselsteijn WA. Personal informatics, self-insight, and behavior change: a critical review of current literature. Hum Comput Interact 2017;32(5-6):268-296 [FREE Full text] [doi: 10.1080/07370024.2016.1276456]

32. Stiglbauer B, Weber S, Batinic B. Does your health really benefit from using a self-tracking device? Evidence from a longitudinal randomized control trial. Comput Hum Behav 2019;94:131-139 [FREE Full text] [doi: 10.1016/j.chb.2019.01.018]

33. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]

34. Vorderstrasse A, Lewinski A, Melkus GD, Johnson C. Social support for diabetes self-management via eHealth interventions. Curr Diab Rep 2016;16(7):56. [doi: 10.1007/s11892-016-0756-0] [Medline: 27155606]

35. Shin G, Jarrahi MH, Fei Y, Karami A, Gafinowitz N, Byun A, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: a systematic literature review. J Biomed Inform 2019;93:103153. [doi: 10.1016/j.jbi.2019.103153] [Medline: 30910623]

36. Alnosayan N, Chatterjee S, Alluhaidan A, Lee E, Houston Feenstra FL. Design and usability of a heart failure mHealth system: a pilot study. JMIR Hum Factors 2017;4(1):e9 [FREE Full text] [doi: 10.2196/humanfactors.6481] [Medline: 28341615]

37. Kelley C, Lee B, Wilcox L. Self-tracking for mental wellness: understanding expert perspectives and student experiences. Proc SIGCHI Conf Hum Factor Comput Syst 2017;2017:629-641 [FREE Full text] [doi: 10.1145/3025453.3025750] [Medline: 28920106]

38. O'Connor S, Hanlon P, O'Donnell C, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]

XSL•FO

RenderX

39.   Karapanos E, Gouveia R, Hassenzahl M, Forlizzi J. Wellbeing in the making: peoples' experiences with wearable activity trackers. Psychol Well Being 2016;6:4 [FREE Full text] [doi: 10.1186/s13612-016-0042-6] [Medline: 27376017]

40.   Pacanowski CR, Linde JA, Neumark-Sztainer D. Self-weighing: helpful or harmful for psychological well-being? A review of the literature. Curr Obes Rep 2015;4(1):65-72 [FREE Full text] [doi: 10.1007/s13679-015-0142-2] [Medline: 26627092]

41.   Etkin J. The hidden cost of personal quantification. J Consum Res 2016;42(6):967-984 [FREE Full text] [doi: 10.1093/jcr/ucv095]

42.   Andersen TO, Langstrup H, Lomborg S. Experiences with wearable activity data during self-care by chronic heart patients: qualitative study. J Med Internet Res 2020;22(7):e15873 [FREE Full text] [doi: 10.2196/15873] [Medline: 32706663]

43.   Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]

Abbreviations

PLR: participant-led research
SUS: System Usability Scale

XSL•FO
RenderX

<u>Original Paper</u>

Utilization of a Smart Sock for the Remote Monitoring of Patients With Peripheral Neuropathy: Cross-sectional Study of a Real-world Registry

Henk Jan Scholten[1*], BA, LLM; Chia-Ding Shih[2], MPH, DPM; Ran Ma[1], BSc; Kara Malhotra[1], MSc; Alexander M Reyzelman[2*], DPM

[1]Siren Care, Inc, San Francisco, CA, United States

[2]California School of Podiatric Medicine, Samuel Merritt University, San Francisco, CA, United States

[*]these authors contributed equally

Corresponding Author:

Henk Jan Scholten, BA, LLM
Siren Care, Inc
1256 Folsom Street
San Francisco, CA, 94103
United States
Phone: 1 6284449603
Email: henkjan.scholten@siren.care

Abstract

Background: Remote patient monitoring (RPM) devices are increasingly being used in caring for patients to reduce risks of complications. Temperature monitoring specifically has been shown in previous studies to provide a useful signal of inflammation that may help prevent foot ulcers.

Objective: In this cross-sectional study, we evaluated utilization data for patients who were prescribed smart socks as remote temperature monitoring devices.

Methods: This study evaluated data from a patient registry from January to July 2021. The utilization data, which were collected starting from the first full month since patients were prescribed the smart socks, were evaluated along with retention over time, the average time that the socks were worn, and the number of days that the socks were worn per month and per week.

Results: A total of 160 patients wore the smart sock RPM device for 22 to 25 days per month on average. The retention rate was 91.9% (147/160) at the end of the 7-month period; a total of 13 patients were lost to follow-up during this period. The average number of days that the socks were worn per week was 5.8. The percentage of patients with a utilization rate of >15 days ranged from 79.7% (106/133) to 91.9% (125/136) each month.

Conclusions: This study shows a high level of utilization for a smart sock RPM device and a high compliance rate. A future prospective study on the clinical outcomes after the use of the smart socks may further solidify the idea of conducting temperature monitoring for foot ulcer prevention.

Introduction

Remote patient monitoring (RPM) has emerged as a critical method in disease prevention. A wide range of devices are designed to monitor physiologic indicators of clinical interest for a variety of health conditions, especially diabetes and related complications. Studies have demonstrated the efficacy and promise of RPM in diabetes management. Su et al [1] examined glycemic control for 1354 patients and concluded that patients who underwent more frequent and regular remote monitoring had lower hemoglobin A_{1c} levels than those of patients with lower adherence to the program. Further, the willingness to adopt remote monitoring was evaluated in 1577 patients with diabetes, and perceived intrusiveness was a main factor for

XSL·FO
RenderX

whether a patient would adopt monitoring in diabetes management [2].

Diabetic foot ulcers (DFUs) are a highly prevalent complication for people living with diabetes, who have an estimated 25% lifetime risk of developing DFUs [3]. Temperature was first identified as a predictive factor for ulceration by Benbow et al [4]. Researchers further developed temperature monitoring by measuring multiple sites on each foot to assess temperature differentials that may predict the onset of a neuropathic ulceration [5]. "Areas that are likely to ulcerate have been associated with increased local skin temperatures due to inflammation and enzymatic autolysis of tissue" [6]. Identifying areas of injury through inflammation tracking allows patients and their providers to intervene to reduce inflammation before a wound develops [6].

Various technologies have been developed that use temperature differentials to remotely monitor diabetic foot health [7]. One technology is a smart sock that can be worn by patients and has a regular connection to the cloud for the capture and sharing of temperature data with health care professionals. The socks have temperature sensors embedded inside of the fabric, so that it is soft and comfortable for the user. The patient data are monitored for an elevation in temperature that indicates inflammation—an early sign of wound formation. Although the socks are mainly intended for preventative therapy, this product has also been used to track the course of developing ulcers. Additionally, the socks can be used by people who have undergone an amputation; the socks use an ipsilateral temperature algorithm to detect temperature elevations in 1 foot. Reyzelman et al [8] first evaluated the smart socks in a 35-patient study; patients reported that the socks were easy to use and comfortable, ranking them with a median score of 9 and 10 for comfort and ease of use, respectively, on a 10-point scale. These ratings for ease of use and comfort indicate potentially low intrusiveness and a high willingness to use smart socks as an RPM device.

In this cross-sectional study, we reviewed real-world data from patients using a smart sock temperature monitoring device (Siren Socks; Siren Care) to assess compliance and utilization levels. The purpose of this study is to determine how patients adhere to an RPM program for DFU prevention that involves the use of smart socks.

Methods

Study Design

Patients in the Siren Care registry who were enrolled throughout July 2021 were retrospectively reviewed. This registry is an institutional review board (IRB)–approved protocol (IRB submission title: *Temperature and Activity Data from the "Siren Socks and Foot Monitoring System" – A Multicenter Post Market Registry Study with Retrospective and Prospective Analysis*; WCG IRB study number: 1284366). The qualification for subscribing to the smart socks was the diagnosis of peripheral neuropathy. The inclusion criteria were patients using the smart sock temperature monitoring device for at least 1 day, those who were at participating sites, and those who consented to the registry, as per the IRB-approved protocol. The first calendar month of utilization was excluded due to the variability in which days of the month patients began using the RPM device. The utilization data were measured as the amount of time that the socks were worn, as measured by the smart sock device, in terms of wear time per day as well as the number of days that the socks were worn per month and per week.

Description of the Temperature Monitoring Device

The smart sock device takes continuous measurements of temperature at 6 points on each foot (the hallux, the heel, the arch, metatarsal 1, metatarsal 3, and metatarsal 5). The temperature is measured automatically throughout the day. The socks turn on upon wear and turn off automatically when they are no longer worn. No charging is necessary by the patient. No smartphone is necessary to be used by the patient for data transmission. A hub is plugged into the wall for data transmission, and monitoring data are also stored on the socks to allow for monitoring when patients are away from home. The socks are designed to be machine washable and can be reused for a period of up to 1 year (Figure 1). The socks' lifetime is around 1 year, but this can be longer depending on usage. Socks are replaced after normal wear and tear. Data were collected and monitored by prescribing physicians and their designated staff. Any temperature differentials greater than 2.2 °C resulted in an alert to a monitoring nurse that required follow-up via a phone call to the patient.

Figure 1. The smart sock remote temperature monitoring device (Siren Socks; Siren Care).
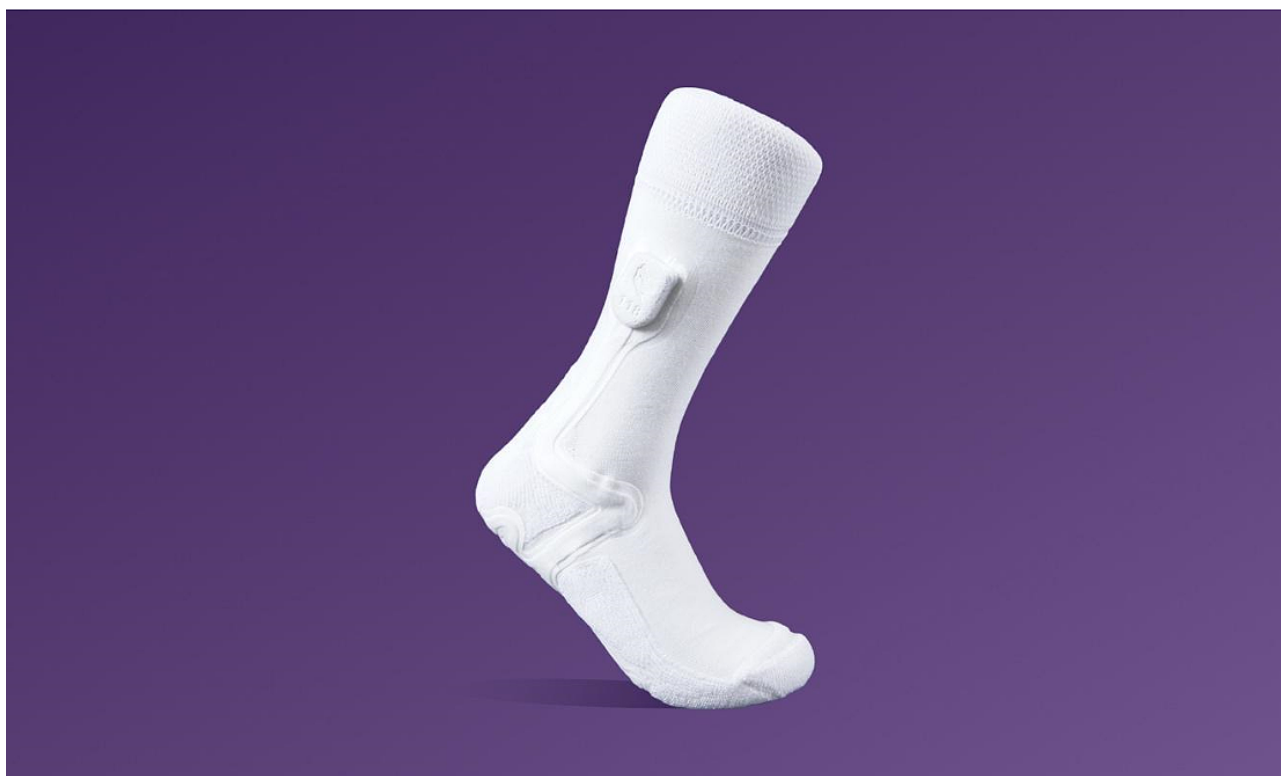
Description of Outcome Measures

The primary outcome measure was the number of days that the socks were worn per month and per week. The secondary outcome measure was the amount of time that the socks were worn per day.

Statistical Analysis

Descriptive statistics were performed, and the analysis was performed by using Microsoft Excel (version 16.51; Microsoft Corporation).

Ethics Approval

This study was approved by the WCG IRB (study number: 12843666). If an individual wished to participate in the study, they were informed about the study objectives, and they could consent through a mobile app or over the phone after having started using the socks.

Results

A total of 160 patients met the inclusion criteria as of July 2021. Data from previous months of enrollment were included, beginning in January 2021 (Table 1).

Table 1. Smart sock utilization (>15 days and >5 days of wear).

Utilization characteristics	January 2021	February 2021	March 2021	April 2021	May 2021	June 2021	July 2021
Active patients, n	10	20	23	62	110	133	136
Patients with >15 days of wear, n (%)	8 (80)	17 (85)	20 (87)	55 (89)	95 (86)	106 (80)	125 (92)
Patients with >5 days of wear, n (%)	10 (100)	20 (100)	22 (96)	60 (97)	108 (98)	125 (94)	132 (97)

Across the entire studied period, patients wore the socks for 5.8 (SD 1.7) days per week on average. The median number of days that the socks were worn was 7 days per week. Further, 93% of the time, socks were worn for at least 3 days per week (Table 2).

XSL•FO
RenderX

Table 2. Smart sock utilization (number of days worn per month and week and number of hours worn per day).

Utilization characteristics	January 2021	February 2021	March 2021	April 2021	May 2021	June 2021	July 2021
Number of wear days per month, mean (SD; median)	23.0 (7.8; 26)	23.1 (7.8; 27)	24.8 (8.3; 28)	23.7 (7.2; 26)	23.9 (7.4; 26.5)	22.0 (8.3; 24)	25.6 (7.4; 29)
Number of wear days per week, mean (SD; median)	5.5 (2.0; 6.5)	6.1 (1.5; 7)	5.8 (1.7; 7)	5.9 (1.5; 7)	5.7 (1.7; 6)	5.7 (1.7; 6)	6.0 (1.6; 7)
Number of wear hours per day, mean (SD; median)	12.4 (3.1; 12.5)	13.8 (4.6; 13.7)	12.6 (5.2; 13.0)	11.7 (4.4; 11.8)	11.0 (4.7; 10.6)	11.0 (4.9; 11.4)	10.0 (5.2; 9.7)

The wear time per day was assessed across all of the months that the socks were worn. The average wear time per day was 11.0 (SD 4.9) hours, and the median was 11.1 hours (Table 2).

In total, 24 people ceased to be active during the studied period; 11 patients temporarily paused their use of the socks, and 13 were lost to follow-up and went off service (Tables 3 and 4).

Table 3. Smart sock retention.

Retention characteristics	January 2021	February 2021	March 2021	April 2021	May 2021	June 2021	July 2021
Active patients, n	10	20	23	62	110	133	136
Patients retained (ie, those still active at the end of July), n (%)	9 (90)	17 (85)	20 (87)	52 (84)	95 (86)	122 (92)	N/A[a]

[a]N/A: not applicable.

Table 4. Smart sock retention: reasons for inactivity.

Reason	Patients, n
Paused temporarily due to other health conditions (eg, healing from surgery, open wounds, or other health issues)	6
Paused while resolving technical issues	3
Paused while waiting for cooler weather	2
Deceased	1
Changed providers when moving into a permanent nursing home	1
Returns related to comfort	2
Returns due to a lack of education on the use and intent of the device	4
Lost to follow-up and did not respond to repeated calls	5

Patients' average age at the time of enrollment was 69.9 (SD 10.7; median 71) years. The youngest patient was 37 years old, and the oldest was 94 years old.

Discussion

Principal Findings

This study appears to be the first to use data from patients who were being tracked over time during normal pediatric practice to analyze adherence to and compliance with an RPM device for plantar foot temperature. Overall, the temperature monitoring smart socks offered patients with peripheral neuropathy a reliable, easy-to-comply, and real-time device designed to help reduce the risk of foot ulceration. Our findings indicated a high utilization rate of 22 to 25 days per month and a retention rate of 91.9% (147/160; Table 1). The information transmitted from the temperature monitoring smart socks thus allowed providers to closely monitor these high-risk patients.

A number of novel RPM technologies have been developed to provide patients and clinicians with options for monitoring temperature, as temperature is a physiological indicator of inflammation and possibly an early warning sign of foot ulcer formation [4-6]. However, these RPM technologies' promise for preventing foot ulcers is based on patients' ability and willingness to use such devices in their activities of daily living. Patients' compliance with the use of medical devices is known to be poor when it is burdensome to their daily life. A prior study by Armstrong et al [9] found that the utilization rate of using a removable cast walker as an offloading device for DFUs was low [10]. The compliance rate for the offloading device was only as high as 28%. The compliance and utilization rates for the smart socks were considerably higher, despite the average age of the enrollees at the time of enrollment being 69.9 (SD 10.7) years. This suggests that the technology is easy to use, even for older users, who are at increased risk of diabetes and DFUs.

Smart socks, smart pads, and smart insoles are among the RPM devices discussed in the literature [9,11,12]. The reported utilization rate of smart insoles is roughly 6.1 to 6.9 hours per day, and that of the smart pad averages from 1.6 to 4.1 days per week [13,14]. The smart socks in this study reported a high utilization and compliance rate; the socks were used for 22 to 25 days per month and 5.8 days per week on average (Table 2).

The results of this study suggest that the smart sock was used by patients to a high degree, as patients wore the device for an average of 22 to 25 days per month during the period studied. Notably, the percentage of patients that wore the device for at least 15 days in a month ranged from 79.7% (106/133) to 91.9% (125/136) for any particular month (Table 1). The number of patients who wore the RPM device for >5 days per month was high, ranging from 94% (125/133) to 100% (20/20) each month (Table 1). This high number of patients suggests that many patients who fail to achieve >15 days of wear time can be guided to increase their frequency of wear. The wear time per day was, on average, 11.1 hours (Table 2). This suggests that the smart socks were not worn for a brief period but rather were worn extensively throughout the day. Given our findings, the smart socks achieved a high compliance, utilization, and retention rate.

The retention rate was analyzed to be 91.9% (147/160), with only 13 patients dropping out by going off service either through returns or by being lost to follow-up. Further, 11 patients were still on service, but they temporarily paused their use of the socks due to comorbidities or technical issues (Table 4). A total of 160 patients were enrolled, and 149 were still on service by the end of the study period. Patient retention was reviewed by month, and many of the patients in the registry were added in the middle of the study period.

No specific user research was done into the reasons for the high compliance and utilization rates, but a possible reason for these may be that socks are a simple and unintrusive form for an RPM device. Additionally, the lack of charging and regular contact with a nurse for assistance with the RPM services may have also contributed to the high level of utilization. Further analysis, perhaps through patient questionnaires, may provide further insight into the reasons for high utilization. Self-management and the actual utilization of preventative services and devices are important factors for determining health outcomes in chronic conditions. In general, compliance is defined as "the extent to which a person's behavior coincides with medical advice" [15]. The International Working Group on the Diabetic Foot released guidelines in 2019 that included a recommendation for temperature monitoring and a daily self-inspection of feet for patients at risk of ulceration [16]. The level of compliance to foot care advice has been studied to a limited degree. One study found that only 38.7% of a sample of 331 patients examined their feet 5 to 7 days per week [17]. Adherence to recommendations for foot temperature monitoring has not been extensively studied. One study did demonstrate that a 50% rate of adherence to recording foot temperature resulted in a significantly lower likelihood of developing an ulcer when compared with lower rates of adherence [18]. These findings suggest that adherence may be a challenge with regard to self-management behaviors among patients with diabetes and that adherence is a meaningful factor. In our study, based on the early results of the utilization of the smart socks, patients have a high level of adherence to prescribed advice on wearing smart socks in a real-world setting.

This study also has a few limitations. The period of observation was limited to 7 months, and many patients entered the registry during the middle and later parts of the evaluation period. Further follow-up and a greater number of patients would be necessary to better assess changes in utilization and retention over time.

Conclusion

The usefulness of temperature monitoring for podiatric patients with limited or no protective sensation has been demonstrated [5,19,20]. The level of adherence to and the utilization of various temperature monitoring devices need further evaluation. This study shows a high level of utilization and compliance for a smart sock remote temperature monitoring device. Further studies with larger patient groups and a longer follow-up period are warranted to better understand the sustained adherence to RPM among patients with diabetes.

References

1. Su D, Michaud TL, Estabrooks P, Schwab RJ, Eiland LA, Hansen G, et al. Diabetes management through remote patient monitoring: The importance of patient activation and engagement with the technology. Telemed J E Health 2019 Oct;25(10):952-959. [doi: 10.1089/tmj.2018.0205] [Medline: 30372366]
2. Oikonomidi T, Ravaud P, Cosson E, Montori V, Tran VT. Evaluation of patient willingness to adopt remote digital monitoring for diabetes management. JAMA Netw Open 2021 Jan 04;4(1):e2033115 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.33115] [Medline: 33439263]
3. Singh N, Armstrong DG, Lipsky BA. Preventing foot ulcers in patients with diabetes. JAMA 2005 Jan 12;293(2):217-228. [doi: 10.1001/jama.293.2.217] [Medline: 15644549]
4. Benbow SJ, Chan AW, Bowsher DR, Williams G, Macfarlane IA. The prediction of diabetic neuropathic plantar foot ulceration by liquid-crystal contact thermography. Diabetes Care 1994 Aug;17(8):835-839. [doi: 10.2337/diacare.17.8.835] [Medline: 7956627]
5. Armstrong DG, Holtz-Neiderer K, Wendel C, Mohler MJ, Kimbriel HR, Lavery LA. Skin temperature monitoring reduces the risk for diabetic foot ulceration in high-risk patients. Am J Med 2007 Dec;120(12):1042-1046. [doi: 10.1016/j.amjmed.2007.06.028] [Medline: 18060924]

6.   Lavery LA, Higgins KR, Lanctot DR, Constantinides GP, Zamorano RG, Armstrong DG, et al. Home monitoring of foot
     skin temperatures to prevent ulceration. Diabetes Care 2004 Nov;27(11):2642-2647. [doi: 10.2337/diacare.27.11.2642]
     [Medline: 15504999]

7.   Golledge J, Fernando M, Lazzarini P, Najafi B, Armstrong DG. The potential role of sensors, wearables and telehealth in
     the remote management of diabetes-related foot disease. Sensors (Basel) 2020 Aug 13;20(16):4527 [FREE Full text] [doi:
     10.3390/s20164527] [Medline: 32823514]

8.   Reyzelman AM, Koelewyn K, Murphy M, Shen X, Yu E, Pillai R, et al. Continuous temperature-monitoring socks for
     home use in patients with diabetes: Observational study. J Med Internet Res 2018 Dec 17;20(12):e12460 [FREE Full text]
     [doi: 10.2196/12460] [Medline: 30559091]

9.   Armstrong DG, Lavery LA, Kimbriel HR, Nixon BP, Boulton AJM. Activity patterns of patients with diabetic foot ulceration:
     patients with active ulceration may not adhere to a standard pressure off-loading regimen. Diabetes Care 2003
     Sep;26(9):2595-2597. [doi: 10.2337/diacare.26.9.2595] [Medline: 12941724]

10. Gordon IL, Rothenberg GM, Lepow BD, Petersen BJ, Linders DR, Bloom JD, et al. Accuracy of a foot temperature
     monitoring mat for predicting diabetic foot ulcers in patients with recent wounds or partial foot amputation. Diabetes Res
     Clin Pract 2020 Mar;161:108074 [FREE Full text] [doi: 10.1016/j.diabres.2020.108074] [Medline: 32109516]

11. Najafi B, Mohseni H, Grewal GS, Talal TK, Menzies RA, Armstrong DG. An optical-fiber-based smart textile (Smart
     Socks) to manage biomechanical risk factors associated with diabetic foot amputation. J Diabetes Sci Technol 2017
     Jul;11(4):668-677 [FREE Full text] [doi: 10.1177/1932296817709022] [Medline: 28513212]

12. Najafi B, Ron E, Enriquez A, Marin I, Razjouyan J, Armstrong DG. Smarter sole survival: Will neuropathic patients at
     high risk for ulceration use a smart insole-based foot protection system? J Diabetes Sci Technol 2017 Jul;11(4):702-713
     [FREE Full text] [doi: 10.1177/1932296816689105] [Medline: 28627227]

13. Abbott CA, Chatwin KE, Foden P, Hasan AN, Sange C, Rajbhandari SM, et al. Innovative intelligent insole system reduces
     diabetic foot ulcer recurrence at plantar sites: a prospective, randomised, proof-of-concept study. Lancet Digit Health 2019
     Oct;1(6):e308-e318 [FREE Full text] [doi: 10.1016/S2589-7500(19)30128-1] [Medline: 33323253]

14. Isaac AL, Swartz TD, Miller ML, Short DJ, Wilson EA, Chaffo JL, et al. Lower resource utilization for patients with healed
     diabetic foot ulcers during participation in a prevention program with foot temperature monitoring. BMJ Open Diabetes
     Res Care 2020 Oct;8(1):e001440 [FREE Full text] [doi: 10.1136/bmjdrc-2020-001440] [Medline: 33055233]

15. Haynes RB, Sackett D, Taylor DW. Compliance in Health Care. Baltimore, Maryland: The Johns Hopkins University Press;
     1979.

16. Bus SA, Lavery LA, Monteiro-Soares M, Rasmussen A, Raspovic A, Sacco ICN, International Working Group on the
     Diabetic Foot. Guidelines on the prevention of foot ulcers in persons with diabetes (IWGDF 2019 update). Diabetes Metab
     Res Rev 2020 Mar;36 Suppl 1:e3269. [doi: 10.1002/dmrr.3269] [Medline: 32176451]

17. Neta DSR, da Silva ARV, da Silva GRF. Adherence to foot self-care in diabetes mellitus patients. Rev Bras Enferm
     2015;68(1):103-108, 111-116 [FREE Full text] [doi: 10.1590/0034-7167.2015680115p] [Medline: 25946502]

18. Lavery LA, Higgins KR, Lanctot DR, Constantinides GP, Zamorano RG, Athanasiou KA, et al. Preventing diabetic foot
     ulcer recurrence in high-risk patients: use of temperature monitoring as a self-assessment tool. Diabetes Care 2007
     Jan;30(1):14-20. [doi: 10.2337/dc06-1600] [Medline: 17192326]

19. Yavuz M, Ersen A, Hartos J, Lavery LA, Wukich DK, Hirschman GB, et al. Temperature as a causative factor in diabetic
     foot ulcers: A call to revisit ulceration pathomechanics. J Am Podiatr Med Assoc 2019 Sep;109(5):345-350 [FREE Full
     text] [doi: 10.7547/17-131] [Medline: 30427732]

20. Frykberg RG, Gordon IL, Reyzelman AM, Cazzell SM, Fitzgerald RH, Rothenberg GM, et al. Feasibility and efficacy of
     a smart mat technology to predict development of diabetic plantar ulcers. Diabetes Care 2017 Jul;40(7):973-980. [doi:
     10.2337/dc16-2294] [Medline: 28465454]

Abbreviations

DFU: diabetic foot ulcer
IRB: institutional review board
RPM: remote patient monitoring

XSL•FO

RenderX

Original Paper

Performance of a Computational Phenotyping Algorithm for Sarcoidosis Using Diagnostic Codes in Electronic Medical Records: Case Validation Study From 2 Veterans Affairs Medical Centers

Mohamed I Seedahmed[1,2], MPH, MD; Izabella Mogilnicka[2,3], MD; Siyang Zeng[2,4], MS; Gang Luo[4], PhD; Mary A Whooley[2,5,6], MD; Charles E McCulloch[7], PhD; Laura Koth[1*], MD; Mehrdad Arjomandi[1,2*], MD

[1]Division of Pulmonary, Critical Care, Allergy and Immunology, and Sleep, Department of Medicine, University of California San Francisco, San Francisco, CA, United States

[2]San Francisco Veterans Affairs Medical Center, San Francisco, CA, United States

[3]Department of Experimental Physiology and Pathophysiology, Laboratory of the Centre for Preclinical Research, Medical University of Warsaw, Warsaw, Poland

[4]Department of Biomedical Informatics and Medical Education, School of Medicine, University of Washington, Seattle, WA, United States

[5]Department of Medicine, University of California San Francisco, San Francisco, CA, United States

[6]Measurement Science Quality Enhancement Research Initiative, San Francisco Veterans Affairs Healthcare System, San Francisco, CA, United States

[7]Department of Epidemiology & Biostatistics, University of California San Francisco, San Francisco, CA, United States

[*]these authors contributed equally

Corresponding Author:
Mohamed I Seedahmed, MPH, MD
Division of Pulmonary, Critical Care, Allergy and Immunology, and Sleep
Department of Medicine
University of California San Francisco
513 Parnassus Ave
HSE 1314, Box 0111
San Francisco, CA, 94143
United States
Phone: 1 (415) 476 0735
Fax: 1 (415) 502 2605
Email: mohamed.seedahmed@ucsf.edu

Abstract

Background: Electronic medical records (EMRs) offer the promise of computationally identifying sarcoidosis cases. However, the accuracy of identifying these cases in the EMR is unknown.

Objective: The aim of this study is to determine the statistical performance of using the International Classification of Diseases (ICD) diagnostic codes to identify patients with sarcoidosis in the EMR.

Methods: We used the ICD diagnostic codes to identify sarcoidosis cases by searching the EMRs of the San Francisco and Palo Alto Veterans Affairs medical centers and randomly selecting 200 patients. To improve the diagnostic accuracy of the computational algorithm in cases where histopathological data are unavailable, we developed an *index of suspicion* to identify cases with a *high index of suspicion* for sarcoidosis (confirmed and probable) based on clinical and radiographic features alone using the American Thoracic Society practice guideline. Through medical record review, we determined the positive predictive value (PPV) of diagnosing sarcoidosis by two computational methods: using ICD codes alone and using ICD codes plus the *high index of suspicion*.

Results: Among the 200 patients, 158 (79%) had a high index of suspicion for sarcoidosis. Of these 158 patients, 142 (89.9%) had documentation of nonnecrotizing granuloma, confirming biopsy-proven sarcoidosis. The PPV of using ICD codes alone was 79% (95% CI 78.6%-80.5%) for identifying sarcoidosis cases and 71% (95% CI 64.7%-77.3%) for identifying histopathologically confirmed sarcoidosis in the EMRs. The inclusion of the generated *high index of suspicion* to identify confirmed sarcoidosis cases

XSL•FO
RenderX

increased the PPV significantly to 100% (95% CI 96.5%-100%). Histopathology documentation alone was 90% sensitive compared with *high index of suspicion*.

Conclusions: ICD codes are reasonable classifiers for identifying sarcoidosis cases within EMRs with a PPV of 79%. Using a computational algorithm to capture *index of suspicion* data elements could significantly improve the case-identification accuracy.

Introduction

Background

Sarcoidosis is a complex disease with an unknown etiology that can involve multiple organs, and no universal or standardized measures can fully secure its final diagnosis [1-3]. In fact, it was only recently that the American Thoracic Society (ATS) published its first practice guideline to provide recommendations for diagnosing sarcoidosis and the necessary screening tests [3]. The ATS practice guideline for diagnosis requires the presence of specific clinical and radiographic features, tissue biopsy revealing nonnecrotizing granulomas, and exclusion of alternative conditions that can mimic sarcoidosis [1,3,4].

Data from electronic medical records (EMRs) are commonly used in research and by health care systems, including the United States Department of Veterans Affairs (VA), to predict outcomes or assess care quality [5]. EMR data are generally captured in two forms: (1) *structured data*, including billing codes such as the International Classification of Diseases (ICD) codes, laboratory test results, and procedural codes; and (2) *narrative or unstructured data*, including progress notes, pathology reports, and imaging reports. ICD codes cast a wider net to capture patients in the EMR because they include both inpatient and outpatient claims compared with other classifiers such as Diagnosis-Related Group that only capture inpatient claims [6]. Unstructured data contain many more details of the clinical conditions, but extracting these details is challenging and time consuming. In contrast, structured data are easier to search for, and they allow for identifying cases computationally using diagnostic codes. However, diagnostic codes can be inaccurate and difficult to verify. This is particularly true for the case definition of sarcoidosis, which is considered a diagnosis of exclusion and requires a review of clinical, radiological, and histopathological data for accurate diagnosis [3,7,8]. A few studies have reported the development of sarcoidosis-specific "computationally identifying algorithms" based on structured data elements in the EMR, although they were not validated by manual chart review [9-13]. Another study assessed the accuracy of using diagnostic codes to identify sarcoidosis cases [14] but only used the ICD, Ninth Revision (ICD-9) code and not the ICD, Tenth Revision (ICD-10) code, and it did not include any computational algorithm development. In addition, previous studies on the diagnostic accuracy of ICD codes for other common pulmonary diseases that have less or similar complexity compared with sarcoidosis, such as chronic obstructive pulmonary disease, idiopathic pulmonary fibrosis, and asthma, showed positive predictive values (PPVs) of 42%-67% [15-17].

Moreover, researchers have previously developed predictive models and risk scores to use advanced computational methods to predict, commonly, less-complex case definitions in the EMR [18-24]. For example, in a study published by Himes et al [19], Bayesian network machine learning models were constructed to predict chronic obstructive pulmonary disease. Therefore, given the complexity of securing a sarcoidosis diagnosis in the realm of real-world clinical data, it is essential to develop automated algorithms to detect confirmed and probable cases of sarcoidosis using data elements from structured and unstructured domains by incorporating the ATS diagnostic criteria [3,25].

First Step

As the first step in evaluating the knowledge gap in developing future sarcoidosis-specific "computationally identifying algorithms," we designed this study (1) to estimate the statistical performance of using diagnostic codes (ICD-9 and ICD-10) alone compared with a new approach that uses additional information from radiology and clinical domains, but not histopathology, to inform the utility of these codes for performing clinical phenotyping of sarcoidosis cases in large EMR data sets of the VA and (2) to assess the computational challenges in querying sarcoidosis cases and extracting high-quality sarcoidosis-related research variables from the EMR accurately.

Methods

Data Source and Collection

This was an observational retrospective study of EMRs available through VA Informatics and Computing Infrastructure (VINCI). VINCI provides access to comprehensive and integrated veterans' national deidentified data sets and offers the necessary computational and analytical tools in a secure, high-performance computing environment [26,27]. This study was approved by the institutional review board of the University of California San Francisco and the Veterans Health Administration Research and Development Committee (15-16660). Patients or the public were not involved in the design, conduct, reporting, or dissemination plans of our research.

We searched the EMR data in VINCI from 1989 to 2019 and identified all patients coded as having sarcoidosis in the VA health care system, as defined by the documentation of the ICD-9 and ICD-10 codes of 135 and D86.x (including subcodes), respectively. Data were extracted through executing

SQL queries in an SQL Server 2017 database. A total of 14,833 sarcoidosis cases were identified.

Study Design

To determine the statistical performance of using diagnostic codes (ICD-9 and ICD-10) in identifying patients with sarcoidosis from the EMR, initially, we identified patients with at least one claim (inpatient or outpatient) of ICD diagnosis code for sarcoidosis. To ascertain the true diagnosis of sarcoidosis based on the ATS diagnostic criteria (clinical, radiographic, and pathological findings, as well as exclusion of other causes) [3], 2 clinicians (MIS and IM) performed a comprehensive chart review. Of the 14,833 identified cases, a total of 200 (1.35%) were reviewed to limit the required chart review to a manageable level. As our access to the detailed medical records was limited to the two medical centers of San Francisco VA (SFVA) and Palo Alto VA (PAVA), the reviewed charts were selected from these two centers. We stratified the list of sarcoidosis cases from the 2 centers by site and used the *lottery* method to randomly select 100 patients from each site without a replacement [28] (Figure 1).

Figure 1. Strengthening the Reporting of Observational Studies in Epidemiology flowchart. Selection criteria for sarcoidosis cases. ATS: American Thoracic Society; ICD: International Classification of Diseases; PA: Palo Alto; SF: San Francisco; VA: Veterans Affairs.
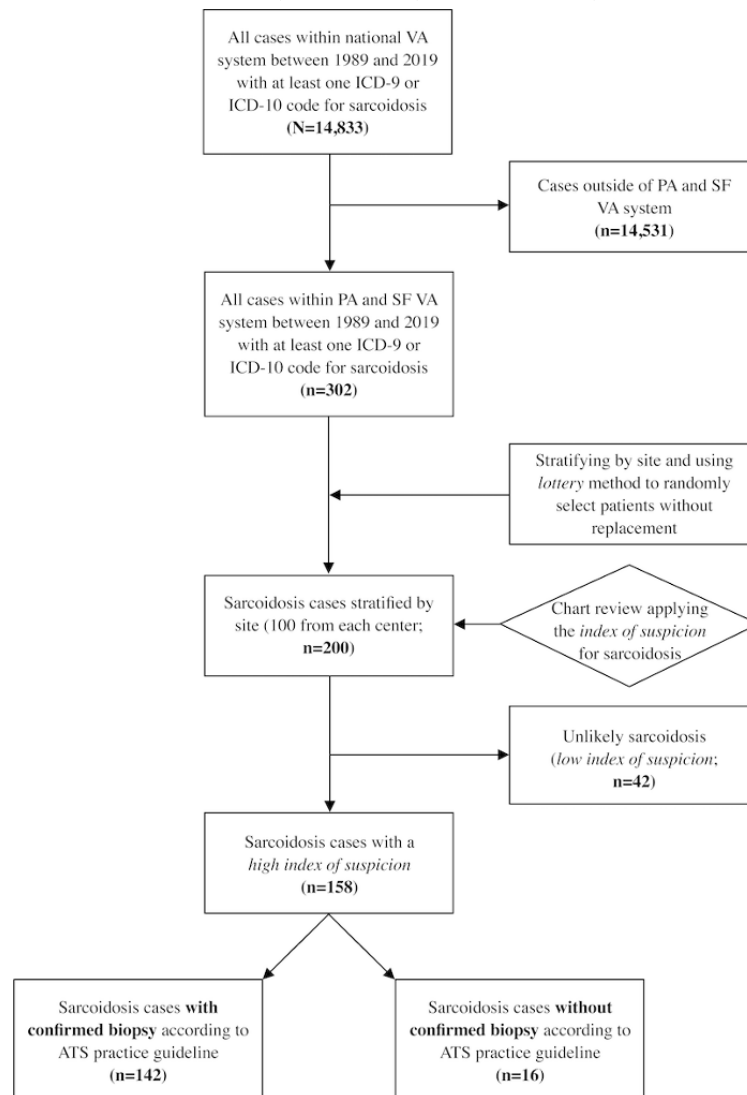
On the basis of the ATS practice guidelines, the diagnosis can be confirmed for those who had a biopsy consistent with sarcoidosis, as well as consistent clinical and radiological findings and no evidence for an alternative diagnosis. However, the ATS practice guideline committee acknowledged that there were clinical situations in which a confirmatory biopsy may not be indicated or possible. Accordingly, based on the ATS practice guideline, those patients without biopsies can be classified as probable sarcoidosis [3]. Therefore, given that not all suspected patients have a tissue biopsy in clinical practice, we generated an *index of suspicion* for sarcoidosis to identify patients with sarcoidosis (confirmed and probable) based on clinical and radiographic information, regardless of the availability of biopsy data, and to assess whether this approach would improve the diagnostic accuracy. The *index of suspicion* was applied to the initial cohort of patients with ICD codes for sarcoidosis (n=200). The clinical and radiological features were extracted from the available structured and unstructured data without including the histopathology results. If the patients were documented to have one or more of these features, they were assigned to the *high index of suspicion* group (Textbox 1); otherwise, the patients were assigned to the *low index of suspicion* group.

Textbox 1. Criteria to determine high index of suspicion.

Clinical and radiological features supportive of the diagnosis of sarcoidosis that were used for the determination of a *high index of suspicion*. Any patients with at least one of these features were included in the *high index of suspicion* group:

- Clinical

    - Lofgren syndrome (defined as erythema nodosum, bilateral hilar lymphadenopathy, and polyarthralgia or polyarthritis)

    - Heerfordt syndrome (defined as facial nerve palsy, parotid gland enlargement, anterior uveitis, and low-grade fever)

    - Lupus pernio or erythema nodosum

    - Maculopapular or erythematous skin lesions or nodules

    - Facial nerve palsy

    - Symmetrical parotid enlargement

    - Optic neuritis, scleritis, uveitis, or retinitis

    - Lacrimal gland swelling

    - Evidence of granulomatous disease on direct laryngoscopy

    - Hepatomegaly or splenomegaly

    - Shortness of breath, dyspnea on exertion, cough, dizziness, or chest pain

    - Pulmonary function test with obstruction, restriction, or low diffusing capacity of the lungs for carbon monoxide

    - Cardiomyopathy, cardiac arrhythmia, or atrioventricular node block

    - Hypercalcemia, hypercalciuria, nephrolithiasis, or abnormal vitamin D levels

    - Elevated angiotensin-converting enzyme inhibitors or soluble interleukin-2 receptors

    - Bronchoalveolar lavage lymphocytosis

- Radiological

    - Bilateral hilar lymphadenopathy (chest radiograph, computed tomography, and positron emission tomography)

    - Computed tomography chest with perilymphatic nodules tracking the peribronchovascular bundle

    - Diffuse infiltrates (chest radiograph, computed tomography, and positron emission tomography) or computed tomography chest or chest radiograph with fibrosis

    - Cardiac magnetic resonance imaging or positron emission tomography–computed tomography consistent with sarcoidosis

    - Enlargement or nodules in liver or spleen (computed tomography, positron emission tomography, or magnetic resonance imaging)

    - Magnetic resonance imaging brain with increased inflammation

    - Extrathoracic enlarged lymph nodes (computed tomography, magnetic resonance imaging, and positron emission tomography)

We then further classified the patients into 3 groups. Patients with a high index of suspicion and documented histopathological evidence of nonnecrotizing granulomas were categorized into the group of *sarcoidosis with confirmed biopsy*. Patients with a high index of suspicion and either no documented biopsy in the EMR or a biopsy showing no histopathological evidence of nonnecrotizing granulomas were categorized into the group of *sarcoidosis without confirmed biopsy* (probable sarcoidosis). Finally, those with a low index of suspicion were categorized into the group of *unlikely sarcoidosis* (Figure 1).

Using the *index of suspicion* restricts the initially developed sarcoidosis cohort to capture those with a *high index of suspicion* for sarcoidosis from whom we identified confirmed cases. As we started with a random sample of those with sarcoidosis diagnostic codes, the further restriction of the sample to those with a *high index of suspicion* was still a random sample of the combination of both ICD codes and the *index of suspicion*. We compared the statistical performance of the two methods (ICD

code alone vs ICD code with *index of suspicion*) to determine whether the use of this *index of suspicion* could improve the PPV of identification of sarcoidosis cases in the EMR.

This approach provides more information than just relying on ICD codes alone to develop robust computational sarcoidosis-specific algorithms consistent with the recent ATS practice guideline recommendations.

Disease-Related Variables

Organ involvement was assessed based on the clinical history obtained from physicians' notes and imaging and biopsy reports available in the computerized patient record system. For this assessment, to adjust for the variability in providers' documentation, we adapted a set of criteria previously introduced in the National Institutes of Health–sponsored Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis (GRADS) study [29].

We collected the following data from the chart review: clinical site, gender, race, ICD-9 and ICD-10 codes for sarcoidosis (135 and D86, respectively), the pathological diagnosis from any available biopsy, organ involvement as described in Textbox 2, Scadding staging of chest x-ray (as described in radiology reports), history of bilateral hilar lymphadenopathy (based on radiology reports and clinical notes), pulmonary function test (PFT) pattern (as reported in PFT reports), the clinical status (acute, chronic, or remitting disease), and the treatment status of sarcoidosis.

Pathological diagnoses were categorized into *primary* histopathological if the data were available in the pathology report domains and *secondary* if the data were available only in the clinical note domains because of either a remote history of biopsy or because the biopsy had been performed outside the VA. The PFT reports at the SFVA and PAVA used Crapo reference equations to calculate the lower limit of normal values for spirometry and lung volume measurements.

Using the clinical data from chart abstraction, we classified the patients into the clinical phenotypes proposed by the GRADS study, with the exception of *multi-organ phenotype*, which we defined as the involvement of ≥3 organs.

Textbox 2. Organ involvement assessment for sarcoidosis (with and without confirmed biopsy).

Organ and assessment

Lung

- Positive lung biopsy and positive mediastinal or hilar lymph node biopsy

- Chest x-ray, computed tomography (CT) chest, or positron emission tomography (PET) demonstrating bilateral hilar lymphadenopathy; CT chest with perilymphatic nodules tracking the peribronchovascular bundle; chest X-ray, CT chest, or PET with diffuse infiltrates; and CT chest or chest x-ray (CXR) with fibrosis

- Pulmonary function test (PFT) with obstruction, restriction, or low diffusing capacity of the lungs for carbon monoxide (DLCO)

Skin

- Positive skin biopsy

- Lupus pernio and erythema nodosum

Eye

- Positive conjunctival or scleral biopsy

- Optic neuritis, scleritis, uveitis, or retinitis

Cardiac

- Positive heart or pericardium biopsy

- Atrioventricular node block (second or third degree)

- Cardiomyopathy responsive to treatment

- Cardiac arrhythmia (eg, ventricular tachycardia)

- Cardiac magnetic resonance imaging (MRI) or PET-CT consistent with sarcoidosis

Liver or spleen

- Positive liver or spleen biopsy

- Enlargement or nodules in liver or spleen (CT, PET, or MRI)

- Abnormal liver enzymes

Neurosarcoidosis

- Positive brain or dura or peripheral nerve biopsy

- Clinical syndrome or symptoms consistent with central nervous system sarcoidosis along with a positive MRI

Ear, nose, and throat

- Positive biopsy from ear, nose, or throat

- Direct laryngoscopy consistent with granulomatous disease

Multi-organ involvement

- ≥3 organs involved based on other criteria in this table

Statistical Analyses

All statistical analyses were performed with R software (The R Foundation for Statistical Computing) using RStudio (version 1.2.5). Descriptive statistics were computed to summarize the data. Categorical variables were presented as the frequency in percentages, and continuous data were presented as means and SDs. We estimated the PPV of the two aforementioned computational diagnostic criteria for sarcoidosis (ICD codes alone and ICD codes along with *index of suspicion*). We did not report the positive likelihood ratio, given that the specificity for using ICD codes alone could not be calculated because our study design did not include a review of noncases. The PPV for the criterion of using only the ICD code was calculated as the number of patients with an ICD code for sarcoidosis divided by the total number of patients verified to have sarcoidosis by chart review (*gold standard*). The PPV for the criterion of using the ICD codes and *index of suspicion* was calculated as the total number of patients with a high index of suspicion divided by the number of patients verified to have sarcoidosis by chart review (*gold standard*). The sensitivity of histopathology reports alone compared with chart review was calculated as the total number of patients with a high index of suspicion and confirmed biopsy divided by the number of patients verified to have sarcoidosis by chart review (*gold standard*). We computed 95% CIs using the exact binomial method. For our estimates, significance was defined as *P*<.05.

Results

Patients' Characteristics

A total of 14,833 patients with at least one ICD-9 or ICD-10 diagnostic code of sarcoidosis were identified. The study cohort included patients identified by the ICD codes of sarcoidosis (n=200). Of the 200 patients, 158 (79%) had a *high index of suspicion* for sarcoidosis based on clinical or radiographic findings. Of these 158 patients, 108 (68.4%) were identified with the ICD-9 code of 135 and 50 (31.6%) with the ICD-10 code of D86, and 142 (89.9%) had confirmed sarcoidosis based on histopathological evidence of nonnecrotizing granuloma and were classified as having *sarcoidosis with confirmed biopsy*; the remaining 16 (10.1%) patients with a *high index of suspicion* did not undergo a biopsy and were classified as having *sarcoidosis without confirmed biopsy* (probable sarcoidosis; Figure 1). No patient had nondiagnostic biopsy results for sarcoidosis.

Table 1 summarizes the demographic data and baseline characteristics of patients with sarcoidosis (with and without confirmed biopsy). Among these patients, 89.9% (142/158) were men and there was a higher representation of African American patients than non-Hispanic White patients (85/158, 53.8%, vs 52/158, 32.9%, respectively). Overall, 90.5% (143/158) had a predominant pulmonary phenotype. Among these, 129 had PFT (36, 27.9%, 28, 21.7%, and 25, 19.4%, with restrictive, obstructive, and mixed patterns, respectively) and most were in Scadding stage II (47/143, 32.9%), followed by stage 0 and stage I (27/143, 18.9%, and 26/143, 18.2%, respectively). There was no significant difference in age between those who had a biopsy performed to diagnose sarcoidosis and those who did not (mean 65.5, SD 10.8, years vs mean 69.3, SD 10.3, years, respectively; *P*=.18). In terms of clinical phenotypes, 37.9% (60/158) had a *multi-organ* disease (≥3 organs; there were none with involvement of ≥5 organs), followed by stage II or stage III treated (45/158, 28.5%). Our study cohort did not include any individuals with acute presentation (acute, untreated). Some patients overlapped with multiple clinical groups.

Table 1. Distribution of characteristics and clinical phenotype groups of patients with sarcoidosis (with and without confirmed biopsy; N=158).

Characteristics	Sarcoidosis with confirmed biopsy (n=142), n (%)	Sarcoidosis without confirmed biopsy (n=16)[a], n (%)	P value
Age (years), mean (SD)	65.5 (10.8)	69.3 (10.3)	.18
Sex			.59[b]
Male	127 (89.4)	15 (93.7)	
Female	15 (10.6)	1 (6.3)	
Race			.62[c]
African American	74 (52.1)	11 (68.8)	
Non-Hispanic White	49 (34.5)	3 (18.8)	
Hispanic White	3 (2.1)	0 (0)	
Unknown	12 (8.5)	2 (12.5)	
Other	4 (2.8)	0 (0)	
International Classification of Diseases codes for sarcoidosis			.60[b]
International Classification of Diseases, Ninth Revision	98 (69)	10 (62.5)	
International Classification of Diseases, Tenth Revision	44 (30.9)	6 (37.5)	
Organ involvement			.38[c]
Lung	86 (60.6)	12 (75)	
Multi-organ (pulmonary without cardiac)	39 (27.5)	2 (12.5)	
Multi-organ (pulmonary and cardiac)	4 (2.8)	0 (0)	
Multi-organ (cardiac without pulmonary)	2 (1.4)	0 (0)	
Multi-organ (neither cardiac nor pulmonary)	11 (7.7)	2 (12.5)	
Pulmonary function test pattern[d]			.03[c]
Obstructive	27 (19)	1 (6.3)	
Restrictive	30 (21.1)	6 (37.5)	
Mixed	20 (11.9)	5 (31.3)	
Normal	39 (27.5)	1 (6.3)	
Missing	26 (18.3)	3 (18.8)	
Scadding stage[e]			.06[c]
Stage 0	22 (15.5)	5 (31.3)	
Stage I	23 (16.2)	3 (18.8)	
Stage II	45 (31.7)	2 (12.5)	
Stage III	17 (11.9)	4 (25)	
Stage IV	22 (15.5)	0 (0)	
Missing	13 (9.2)	2 (12.5)	
Clinical phenotype group[f]			.06[c]
Group 1: multi-organ	56 (39.4)	4 (25)	
Group 2: nonacute, stage I, untreated	6 (4.2)	2 (12.5)	
Group 3: stages II-III, treated	42 (29.6)	3 (18.8)	
Group 4: stages II-III, untreated	14 (9.9)	2 (12.5)	
Group 5: stage IV, treated	17 (11.9)	0 (0)	
Group 6: stage IV, untreated	4 (2.8)	2 (12.5)	
Group 7: acute sarcoidosis, untreated	0 (0)	0 (0)	

Characteristics	Sarcoidosis with confirmed biopsy (n=142), n (%)	Sarcoidosis without confirmed biopsy (n=16)[a], n (%)	P value
Group 8: remitting, untreated	30 (21.1)	5 (31.3)	
Group 9: cardiac sarcoidosis, treated	6 (4.2)	0 (0)	

[a]Probable sarcoidosis: cases with clinical and radiological features consistent with sarcoidosis and do not have confirmatory biopsies.

[b]Chi-square test.

[c]Fisher exact test.

[d]Evaluated based on pulmonary function test reports available in the computerized patient record system.

[e]Scored based on reviewers' interpretation of imaging reports using Scadding staging. Stage 0: normal chest radiograph; stage I: hilar or mediastinal nodal enlargement only; stage II: nodal enlargement and parenchymal disease; stage III: parenchymal disease only; stage IV: end-stage lung disease (pulmonary fibrosis).

[f]Clinical phenotype groups [29]: some patients overlapped with multiple clinical groups. Group 1: multi-organ involvement, patients with ≥3 organs involved; group 2: nonacute, stage I, untreated: patients with nonacute sarcoidosis, stage I, never treated for sarcoidosis; group 3: stage II-III, treated: patients with nonacute sarcoidosis, stage II or III, formerly treated for sarcoidosis or treated within 3 months of data review; group 4: stage II-III, untreated: patients with nonacute sarcoidosis, stage II or III, never treated for sarcoidosis; group 5: stage IV, treated: patients with nonacute sarcoidosis, stage IV, formerly treated for sarcoidosis or treated within 3 months of data review; group 6: stage IV, untreated: patients with nonacute sarcoidosis, stage IV, never treated for sarcoidosis; group 7: acute sarcoidosis, untreated: patients with acute sarcoidosis (Lofgren syndrome); group 8: remitting, untreated: patients who have had no evidence of active clinical disease for >1 year; group 9: cardiac sarcoidosis, treated: patients with cardiac manifestations of sarcoidosis, formerly treated for sarcoidosis or treated within 3 months of data review.

Diagnostic Accuracy of ICD Codes

We then calculated the PPV using ICD codes to identify VA patients who met the ATS definition of sarcoidosis from the VINCI database. For this calculation, we used the curated data set of 200 patients. The PPV of using only ICD codes was 79% (95% CI 78.6%-80.5%) for identifying sarcoidosis cases and 71% (95% CI 64.7%-77.3%) for identifying histopathologically confirmed sarcoidosis in the EMR. After chart review, the inclusion of the generated *high index of suspicion* to identify confirmed sarcoidosis cases increased the PPV significantly to 100% (95% CI 96.5%-100%) with 90% sensitivity of histopathology reports alone compared with chart review (Table 2).

Table 2. Contingency 2×2 table of using histopathology reports compared with high index of suspicion for sarcoidosis cases identification (N=200).

Among patients with International Classification of Diseases code for sarcoidosis	High index of suspicion[a] (chart review)		
	Yes	No	Total
Histopathology report[b]			
Confirmed sarcoidosis	142[c]	0[d]	142
Not available[e]	16[f]	42[g]	58
Total	158	42	200

[a]High index of suspicion for sarcoidosis based on both clinical and radiographic evidence but not biopsy.

[b]Available biopsies with primary or secondary histopathological reports.

[c]Sarcoidosis group with histopathological evidence of nonnecrotizing granuloma.

[d]No sarcoidosis group because of lack of sufficient clinical and radiological features consistent with sarcoidosis even in the presence of the histopathological evidence of nonnecrotizing granuloma.

[e]No biopsies were ordered or available in the electronic medical record.

[f]Probable sarcoidosis group without histopathological evidence of nonnecrotizing granuloma.

[g]No sarcoidosis group because of lack of sufficient clinical and radiological features consistent with sarcoidosis, in addition to the absence of the histopathological evidence of nonnecrotizing granuloma.

Discussion

Principal Findings

In this observational retrospective study of VA EMRs, we reviewed the medical records of 200 randomly selected patients with ICD diagnostic codes for sarcoidosis from the SFVA and PAVA medical centers (Figure 1). In this sample, we found that ICD diagnostic codes performed reasonably well with a PPV of 79% for detecting patients with sarcoidosis and 71% for detecting patients with histopathologically confirmed cases as defined by the ATS clinical practice guideline. After applying the developed *index of suspicion* to the initial cohort, we also demonstrated that including a *high index of suspicion* that incorporated information from radiology and clinical domains, but not histopathology, significantly increased the diagnostic accuracy to 100% (95% CI 96.5%-100%). The results of this study will help researchers and health care systems better understand the accuracy of using diagnostic codes alone versus using ICD codes with a *high index of suspicion* for sarcoidosis as classifiers in detecting a complex disease such as sarcoidosis in the EMR. Furthermore, the study highlighted other

computational challenges in querying sarcoidosis cases and accurately extracting high-quality sarcoidosis-related research variables from the EMR. This approach could be adapted to develop automated chart review algorithms using additional data elements from structured and unstructured domains by applying advanced computational methodologies such as natural language processing (NLP) and machine learning.

The randomly selected cohort of veterans in this study with sarcoidosis (with and without confirmed biopsy) consisted of 89.9% (142/158) of men and 10.1% (16/158) of women. Although the gender distribution in our study was different from that in A Case Control Etiologic Study of Sarcoidosis [30], it is closely reflective of the demographics in the veterans' population [31]. This study confirmed the higher prevalence of sarcoidosis in African American individuals (85/158, 53.8%) compared with non-Hispanic White individuals (52/158, 32.9%), a finding that many other epidemiological studies on sarcoidosis have previously reported [32-36]. At the same time, the study population was racially diverse, highlighting the potential utility of the VA EMRs for studying sarcoidosis in medically underserved populations [37]. In our study, the PPV was reasonable compared with the study conducted by Ungprasert et al [14] for detecting patients with sarcoidosis in the EMR. This difference could be due to not using the ICD-10 code and having a less diverse population (85% White vs 9% Black).

Using ICD codes alone to extract health information is far more convenient than the time-consuming process of manually reviewing narrative data sets in unstructured data. However, using ICD codes to identify sarcoidosis cases in large data sets with thousands of patients poses several practical challenges. First, given the heterogeneity of sarcoidosis, it is challenging to efficiently confirm the presence of the disease. The verification process requires careful analysis of the available narrative data such as progress notes, imaging reports, and pathology reports to establish the case definition based on the sarcoidosis diagnostic criteria [3]. Second, the precise identification of the type of organ involvement through the EMR is a complex process and requires a thorough review of unstructured data. Although there are subcodes for ICD diagnostic codes that aim to capture the involvement of various organs, health care providers may or may not be familiar with these subcodes and may or may not use them correctly.

Moreover, there are no specific ICD codes for classifying the involvement of some organs in sarcoidosis (such as the central nervous system or gastrointestinal tract) [38]. Third, ICD codes do not determine the extent of the disease, such as described by the stages of a chest x-ray [39], because of a lack of ICD codes for different stages of pulmonary sarcoidosis [38]. Analysis of pulmonary features requires a manual review of every patient's radiology reports and cannot be performed using only ICD codes. Finally, ICD codes do not specify the various sarcoidosis presentations such as acute, remitting, or chronic disease [29,38]. Thus, they cannot be used to classify patients into the previously described phenotype groups.

The definition of clinical phenotypes has become an essential goal for the sarcoidosis scientific community because genetic studies have identified different patterns of gene expression

associated with disease severity and disease course [40,41]. In 2015, the National Heart, Lung, and Blood Institute held a workshop to leverage current scientific knowledge and define platforms to address disease disparities, identify high-risk phenotypes, and improve sarcoidosis outcomes [25]. A total of 9 different steps and research strategies were recommended to expand the scope of sarcoidosis research, including EMR-based research, to provide a unified and multidisciplinary approach. Such an approach is expected to bring together stakeholders interested in reducing the burden and severity of sarcoidosis. However, the major barrier in the efficient use of EMR data is the accurate extraction of research-quality variables, case definitions, and outcomes [42]. Thus, the rapid identification of cases and extraction of relevant clinical variables from the EMR using computational phenotype algorithms have emerged as an important next step in EMR-based research. Furthermore, computational phenotype definitions are also essential for conducting pragmatic clinical trials and comparative effectiveness research, increasing the health care system's capacity to effectively deliver precision medicine for patients with sarcoidosis [43].

The two most applied approaches to defining computational phenotypes are (1) a *high-throughput* phenotype algorithm using only structured data (traditionally, the ICD diagnosis codes) and (2) a *low-throughput* phenotype algorithm that accesses structured and unstructured data to develop a sequential flowchart that should end with a case definition. Such a low-throughput approach uses high-performance computational tools such as NLP to process text and extract information using linguistic rules, thereby eliminating the need for a labor-intensive manual review by researchers [7]. Accordingly, this approach is expected to streamline the development of registries and help enrich EMR-based research studies [44]. Our study highlights the need to develop such automated methods to improve the computational case definition of sarcoidosis. Besides, there are other high-quality sarcoidosis-related research variables, including determining the date of the diagnosis, organ involvements, Scadding stages, and the clinical status (acute, chronic, or remitting disease). This approach will assist in automating the extraction of pre-existing or novel clinical phenotypes more precisely and efficiently from the EMR.

Limitations

Our study includes several limitations. First, primary histopathological reports were not available for all the patients. In the cases where the biopsy report was unavailable (either because of a remote history of the biopsy or because the biopsy had been performed outside the VA), we relied on the *secondary* histopathological reports documented in the providers' narrative within the clinical notes. This approach made the diagnosis of sarcoidosis less robust because the confirmatory biopsy reports in these patients could not be directly verified. However, we used the *index of suspicion* approach to define probable sarcoidosis cases regardless of whether a confirmatory biopsy report was available, which is consistent with the diagnostic algorithm recommended by the ATS practice guideline [3]. Second, our definition of *multi-organ phenotype* involved ≥3 organs, instead of ≥5 organs as proposed by the GRADS study [29]. We chose this approach because none of the evaluated

patients were documented to have involvement of ≥5 organs, thus avoiding having no patients with *multi-organ phenotype*. Lack of patients with involvement of ≥5 organs could be due to EMR-related limitations such as missing data and variability in documentation among providers or simply because these patients were cared for at non-VA tertiary medical centers. Third, the generalizability of our findings obtained from VA EMRs to other populations could be limited because the veterans form a special population with a different demographic distribution and exposure from the general population. However, the EMR data of the VA health care system cover >22 million veterans across the United States and >14,000 patients with sarcoidosis ICD diagnosis codes, providing an enormous number of patients to study a rare disease. Moreover, the number of patients whose records were examined in this study was 200, which could be considered a small sample size. However, we analyzed data from nearly two-third of all patients with diagnostic codes for sarcoidosis in the VA health care system across northern California.

Conclusions

Although ICD codes can be used as reasonable classifiers to identify sarcoidosis cases within EMRs with a PPV of 79%, using computational algorithms to extract clinical and radiographic information (*index of suspicion*) from unstructured data could significantly improve the accuracy of case identification. Furthermore, to increase the efficiency of identifying sarcoidosis cases from large health care databases, more studies are required to develop a novel sarcoidosis-specific computational phenotype algorithm using automated emerging methods (such as machine learning and NLP). Moreover, our study sets the stage for promoting research on developing other such algorithms aiming to generate high-quality sarcoidosis-related research variables, such as determining the date of the diagnosis, organ involvements, Scadding stages, and the clinical status (acute, chronic, or remitting disease).

Authors' Contributions

MIS and MA obtained funding. MIS, LK, and MA conceived and designed the study research and developed the study protocol. MIS, IM, SZ, CEM, LK, and MA worked on the methods. MIS, IM, SZ, CEM, LK, and MA analyzed and interpreted the data. MIS and IM wrote the original draft. MIS, IM, SZ, GL, MAW, CEM, LK, and MA reviewed and edited the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Judson MA. Advances in the diagnosis and treatment of sarcoidosis. F1000Prime Rep 2014;6:89 [FREE Full text] [doi: 10.12703/P6-89] [Medline: 25374667]
2. Keller AZ. Anatomic sites, age attributes, and rates of sarcoidosis in U. S. veterans. Am Rev Respir Dis 1973 Apr;107(4):615-620. [doi: 10.1164/arrd.1973.107.4.615] [Medline: 4697670]
3. Crouser ED, Maier LA, Wilson KC, Bonham CA, Morgenthau AS, Patterson KC, et al. Diagnosis and detection of sarcoidosis. An official American thoracic society clinical practice guideline. Am J Respir Crit Care Med 2020 Apr 15;201(8):e26-e51. [doi: 10.1164/rccm.202002-0251st]
4. Judson MA. The diagnosis of sarcoidosis. Curr Opin Pulm Med 2019 Sep;25(5):484-496. [doi: 10.1097/MCP.0000000000000596] [Medline: 31365383]
5. Szeto HC, Coleman R, Gholami P, Hoffman BB, Goldstein MK. Accuracy of computerized outpatient diagnoses in a Veterans Affairs general medicine clinic. Am J Manag Care 2002 Jan;8(1):37-43 [FREE Full text] [Medline: 11814171]
6. Hsia DC, Krushat WM, Fagan AB, Tebbutt JA, Kusserow RP. Accuracy of diagnostic coding for medicare patients under the prospective-payment system. N Engl J Med 1988 Feb 11;318(6):352-355. [doi: 10.1056/nejm198802113180604]
7. Pendergrass SA, Crawford DC. Using electronic health records to generate phenotypes for research. Curr Protoc Hum Genet 2019 Jan;100(1):e80 [FREE Full text] [doi: 10.1002/cphg.80] [Medline: 30516347]
8. Horsky J, Drucker E, Ramelson H. Accuracy and completeness of clinical coding using ICD-10 for ambulatory visits. AMIA Annu Symp Proc 2017;2017:912-920 [FREE Full text] [Medline: 29854158]

XSL•FO
RenderX

9.   Cameli P, Caffarelli C, Refini RM, Bergantini L, d'Alessandro M, Armati M, et al. Hypercalciuria in sarcoidosis: a specific biomarker with clinical utility. Front Med (Lausanne) 2020;7:568020 [FREE Full text] [doi: 10.3389/fmed.2020.568020] [Medline: 33195314]

10.   Sauer WH, Stern BJ, Baughman RP, Culver DA, Royal W. High-risk sarcoidosis. Current concepts and research imperatives. Annals ATS 2017 Dec;14(Supplement_6):S437-S444. [doi: 10.1513/annalsats.201707-566ot]

11.   Boffetta P, Rabkin CS, Gridley G. A cohort study of cancer among sarcoidosis patients. Int J Cancer 2009 Jun 01;124(11):2697-2700 [FREE Full text] [doi: 10.1002/ijc.24261] [Medline: 19230028]

12.   Kraaijvanger R, Janssen Bonás M, Vorselaars AD, Veltkamp M. Biomarkers in the diagnosis and prognosis of sarcoidosis: current use and future prospects. Front Immunol 2020;11:1443 [FREE Full text] [doi: 10.3389/fimmu.2020.01443] [Medline: 32760396]

13.   Baughman RP, Field S, Costabel U, Crystal RG, Culver DA, Drent M, et al. Sarcoidosis in America. Analysis based on health care use. Ann Am Thorac Soc 2016 Aug;13(8):1244-1252. [doi: 10.1513/AnnalsATS.201511-760OC] [Medline: 27509154]

14.   Ungprasert P, Matteson EL, Crowson CS. Accuracy of diagnostic coding for sarcoidosis in electronic databases: a population-based study. Lung 2017 Dec;195(6):713-715 [FREE Full text] [doi: 10.1007/s00408-017-0054-x] [Medline: 28993879]

15.   Lacasse Y, Daigle J, Martin S, Maltais F. Validity of chronic obstructive pulmonary disease diagnoses in a large administrative database. Can Respiratory J 2012;19(2):e5-e9. [doi: 10.1155/2012/260374]

16.   Blais L, Lemière C, Menzies D, Berbiche D. Validity of asthma diagnoses recorded in the medical services database of Quebec. Pharmacoepidemiol Drug Saf 2006 Apr;15(4):245-252. [doi: 10.1002/pds.1202] [Medline: 16374899]

17.   Ley B, Urbania T, Husson G, Vittinghoff E, Brush DR, Eisner MD, et al. Code-based diagnostic algorithms for idiopathic pulmonary fibrosis. Case validation and improvement. Annals ATS 2017 Jun;14(6):880-887. [doi: 10.1513/annalsats.201610-764oc]

18.   Mani S, Chen Y, Elasy T, Clayton W, Denny J. Type 2 diabetes risk forecasting from EMR data using machine learning. AMIA Annu Symp Proc 2012;2012:606-615 [FREE Full text] [Medline: 23304333]

19.   Himes BE, Dai Y, Kohane IS, Weiss ST, Ramoni MF. Prediction of chronic obstructive pulmonary disease (COPD) in asthma patients using electronic medical records. J Am Med Inform Assoc 2009;16(3):371-379 [FREE Full text] [doi: 10.1197/jamia.M2846] [Medline: 19261943]

20.   Romero-Brufau S, Whitford D, Johnson M, Hickman J, Morlan BW, Therneau T, et al. Using machine learning to improve the accuracy of patient deterioration predictions: Mayo Clinic Early Warning Score (MC-EWS). J Am Med Inform Assoc 2021 Jun 12;28(6):1207-1215. [doi: 10.1093/jamia/ocaa347] [Medline: 33638343]

21.   Liao KP, Ananthakrishnan AN, Kumar V, Xia Z, Cagan A, Gainer VS, et al. Methods to develop an electronic medical record phenotype algorithm to compare the risk of coronary artery disease across 3 chronic disease cohorts. PLoS One 2015;10(8):e0136651 [FREE Full text] [doi: 10.1371/journal.pone.0136651] [Medline: 26301417]

22.   Makino M, Yoshimoto R, Ono M, Itoko T, Katsuki T, Koseki A, et al. Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning. Sci Rep 2019 Aug 14;9(1):11862 [FREE Full text] [doi: 10.1038/s41598-019-48263-5] [Medline: 31413285]

23.   Martucci VL, Liu N, Kerchberger VE, Osterman TJ, Torstenson E, Richmond B, et al. A clinical phenotyping algorithm to identify cases of chronic obstructive pulmonary disease in electronic health records. BioRxiv 2019 (forthcoming). [doi: 10.1101/716779]

24.   Chicco D, Jurman G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Med Inform Decis Mak 2020 Feb 03;20(1):16 [FREE Full text] [doi: 10.1186/s12911-020-1023-5] [Medline: 32013925]

25.   Maier LA, Crouser ED, Martin WJ, Eu J. Executive summary of the NHLBI workshop report: leveraging current scientific advancements to understand sarcoidosis variability and improve outcomes. Ann Am Thorac Soc 2017 Dec;14(Supplement_6):S415-S420 [FREE Full text] [doi: 10.1513/AnnalsATS.201707-563OT] [Medline: 29048937]

26.   Velarde KE, Romesser JM, Johnson MR, Clegg DO, Efimova O, Oostema SJ, et al. An initiative using informatics to facilitate clinical research planning and recruitment in the VA health care system. Contemp Clin Trials Commun 2018 Sep;11:107-112 [FREE Full text] [doi: 10.1016/j.conctc.2018.07.001] [Medline: 30035242]

27.   VA Informatics and Computing Infrastructure (VINCI). U.S. Department of Veterans Affairs. URL: https://www.hsrd.research.va.gov/for_researchers/vinci/ [accessed 2022-02-21]

28.   Daniel J. Sampling Essentials: Practical Guidelines for Making Sampling Choices. Thousand Oaks, California, United States: SAGE Publications; 2012.

29.   Moller DR, Koth LL, Maier LA, Morris A, Drake W, Rossman M, et al. Rationale and design of the genomic research in alpha-1 antitrypsin deficiency and sarcoidosis (GRADS) study. Sarcoidosis protocol. Annals ATS 2015 Oct;12(10):1561-1571. [doi: 10.1513/annalsats.201503-172ot]

30.   Baughman RP, Teirstein AS, Judson MA, Rossman MD, Yeager H, Bresnitz EA, Case Control Etiologic Study of Sarcoidosis (ACCESS) research group. Clinical characteristics of patients in a case control study of sarcoidosis. Am J Respir Crit Care Med 2001 Nov 15;164(10 Pt 1):1885-1889. [doi: 10.1164/ajrccm.164.10.2104046] [Medline: 11734441]

XSL·FO

RenderX

31. Department of veterans affairs statistics at a glance. National Center for Veterans Analysis and Statistics. URL: https://www.va.gov/vetdata/docs/Quickfacts/Stats_at_a_glance_4_6_20.PDF [accessed 2022-02-21]

32. Mirsaeidi M, Machado RF, Schraufnagel D, Sweiss NJ, Baughman RP. Racial difference in sarcoidosis mortality in the United States. Chest 2015 Feb;147(2):438-449 [FREE Full text] [doi: 10.1378/chest.14-1120] [Medline: 25188873]

33. Cozier YC, Berman JS, Palmer JR, Boggs DA, Serlin DM, Rosenberg L. Sarcoidosis in black women in the United States: data from the Black Women's Health Study. Chest 2011 Jan;139(1):144-150 [FREE Full text] [doi: 10.1378/chest.10-0413] [Medline: 20595459]

34. Rybicki BA, Major M, Popovich J, Maliarik MJ, Iannuzzi MC. Racial differences in sarcoidosis incidence: a 5-year study in a health maintenance organization. Am J Epidemiol 1997 Feb 01;145(3):234-241. [doi: 10.1093/oxfordjournals.aje.a009096] [Medline: 9012596]

35. Brito-Zerón P, Kostov B, Superville D, Baughman RP, Ramos-Casals M, Autoimmune Big Data Study Group. Geoepidemiological big data approach to sarcoidosis: geographical and ethnic determinants. Clin Exp Rheumatol 2019;37(6):1052-1064 [FREE Full text] [Medline: 31498063]

36. Arkema EV, Cozier Y. Sarcoidosis epidemiology: recent estimates of incidence, prevalence and risk factors. Curr Opin Pulm Med 2020 Sep;26(5):527-534 [FREE Full text] [doi: 10.1097/MCP.0000000000000715] [Medline: 32701677]

37. Population page - racial and ethnic minority veterans. U.S. Department of Veterans Affairs. URL: https://www.va.gov/HEALTHEQUITY/Race_Ethnicity.asp?utm_content&utm_medium=email&utm_name&utm_source=govdelivery&utm_term [accessed 2022-02-21]

38. The ICD-10 Classification of Mental and Behavioural Disorders. Geneva: World Health Organization; 1993.

39. Scadding JG. Prognosis of intrathoracic sarcoidosis in England. A review of 136 cases after five years' observation. Br Med J 1961 Nov 04;2(5261):1165-1172 [FREE Full text] [doi: 10.1136/bmj.2.5261.1165] [Medline: 14497750]

40. Su R, Li MM, Bhakta NR, Solberg OD, Darnell EP, Ramstein J, et al. Longitudinal analysis of sarcoidosis blood transcriptomic signatures and disease outcomes. Eur Respir J 2014 Oct;44(4):985-993 [FREE Full text] [doi: 10.1183/09031936.00039714] [Medline: 25142485]

41. Zhou T, Zhang W, Sweiss NJ, Chen ES, Moller DR, Knox KS, et al. Peripheral blood gene expression as a novel genomic biomarker in complicated sarcoidosis. PLoS One 2012;7(9):e44818 [FREE Full text] [doi: 10.1371/journal.pone.0044818] [Medline: 22984568]

42. Sharma H, Mao C, Zhang Y, Vatani H, Yao L, Zhong Y, et al. Developing a portable natural language processing based phenotyping system. BMC Med Inform Decis Mak 2019 Apr 04;19(Suppl 3):78 [FREE Full text] [doi: 10.1186/s12911-019-0786-z] [Medline: 30943974]

43. Rachel Richesson, Michelle Smerek. Electronic health records-based phenotyping. In: Rethinking Clinical Trials: A Living Textbook of Pragmatic Clinical Trials. Durham, North Carolina, United States: Duke Clinical Research Institute; 2014.

44. Paul DW, Neely N, Clement M, Riley I, Al-Hegelan M, Phelan M, et al. Development and validation of an electronic medical record (EMR)-based computed phenotype of HIV-1 infection. J Am Med Inform Assoc 2018 Feb 01;25(2):150-157 [FREE Full text] [doi: 10.1093/jamia/ocx061] [Medline: 28645207]

Abbreviations

ATS: American Thoracic Society
EMR: electronic medical record
GRADS: Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis
ICD: International Classification of Diseases
NLP: natural language processing
PAVA: Palo Alto Veterans Affairs
PFT: pulmonary function test
PPV: positive predictive value
SFVA: San Francisco Veterans Affairs
VA: United States Department of Veterans Affairs
VINCI: VA Informatics and Computational Infrastructure

XSL•FO
RenderX

<u>Original Paper</u>

Supporting Behavior Change in Sedentary Adults via Real-time Multidimensional Physical Activity Feedback: Mixed Methods Randomized Controlled Trial

Max James Western[1], PhD; Martyn Standage[1], PhD; Oliver James Peacock[1], PhD; Tom Nightingale[2], PhD; Dylan Thompson[1], PhD

[1]Department for Health, University of Bath, Bath, United Kingdom

[2]School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham, Birmingham, United Kingdom

Corresponding Author:
Max James Western, PhD
Department for Health
University of Bath
Claverton Down
Bath, BA2 7AY
United Kingdom
Phone: 44 1225 383732
Email: <u>M.J.Western@bath.ac.uk</u>

Abstract

Background: Increasing physical activity (PA) behavior remains a public health priority, and wearable technology is increasingly being used to support behavior change efforts. Using wearables to capture and provide comprehensive, visually persuasive, multidimensional feedback with real-time support may be a promising way of increasing PA in inactive individuals.

Objective: This study aims to explore whether a 6-week self-monitoring intervention using composite web-based multidimensional PA feedback with real-time daily feedback supports increased PA in adults.

Methods: A 6-week, mixed methods, 2-armed exploratory randomized controlled trial with 6-week follow-up was used, whereby low to moderately active (PA level [PAL] <2.0) adults (mean age 51.3 years, SD 8.4 years; women 28/51, 55%) were randomly assigned to receive the self-monitoring intervention (36/51, 71%) or waiting list control (15/51, 29%). Assessment of PA across multiple health-harnessing PA dimensions (eg, PAL, weekly moderate to vigorous intensity PA, sedentary time, and steps), psychosocial cognitions (eg, behavioral regulation, barrier self-efficacy, and habit strength), and health were made at the prerandomization baseline at 6 and 12 weeks. An exploratory analysis of the mean difference and CIs was conducted using the analysis of covariance model. After the 12-week assessment, intervention participants were interviewed to explore their views on the program.

Results: There were no notable differences in any PA outcome immediately after the intervention; however, at 12 weeks, moderate-to-large effects were observed with a mean difference in PAL of 0.09 (95% CI 0.02-0.15; effect size [Hedges *g*] 0.8), daily moderate-intensity PA of 24 (95% CI 0-45; Hedges *g*=0.6) minutes, weekly moderate-to-vigorous intensity PA of 195 (95% CI 58-331; Hedges *g*=0.8) minutes, and steps of 1545 (95% CI 581-2553; Hedges *g*=0.7). Descriptive analyses suggested that the differences in PA at 12 weeks were more pronounced in women and participants with lower baseline PA levels. Immediately after the intervention, there were favorable differences in autonomous motivation, controlled motivation, perceived competence for PA, and barrier self-efficacy, with the latter sustained at follow-up. Qualitative data implied that the intervention was highly informative for participants and that the real-time feedback element was particularly useful in providing tangible, day-to-day behavioral support.

Conclusions: Using wearable trackers to capture and present sophisticated multidimensional PA feedback combined with discrete real-time support may be a useful way of facilitating changes in behavior. Further investigation into the ways of optimizing the use of wearables in inactive participants and testing the efficacy of this approach via a robust study design is warranted.

Trial Registration: ClinicalTrials.gov NCT02432924; https://clinicaltrials.gov/ct2/show/NCT02432924

XSL•FO
RenderX

Introduction

Background

The health benefits of leading a physically active life are well-established with higher volumes of physical activity (PA), reducing the risk of numerous chronic diseases, mood disorders, and premature mortality [1-3]. In contrast, physical inactivity and prolonged sedentary time have been shown to be independent risk factors for noncommunicable conditions, including type 2 diabetes, cardiovascular disease, cancer, and musculoskeletal disease [4-6]. In addition to health and well-being ramifications, it is estimated that physical inactivity costs US $53.8 billion for health care systems around the world [7]. Collectively, such data stress the need for wide-reaching, cost-effective solutions. The availability, accuracy, and popularity of wearable technology for capturing PA behavior has surged in recent years and presents a potentially useful, affordable, and accessible tool for driving increases in PA levels [8,9]. However, commercial activity monitors are typically marketed at, and used by, young adults who have relatively high baseline PA levels as a means of monitoring exercise performance. Thus, the effectiveness of PA monitoring in inactive populations remains understudied and undetermined [10]. Sophisticated monitoring technology enables personalized motivational and persuasive feedback for individuals who would benefit from an increase in PA [11].

There are multiple dimensions of PA behavior that can independently affect health and well-being [12,13]. Analysis of wearable-derived data shows that individuals can score high and low on any number of these health-harnessing dimensions such as sedentary time, moderate to vigorous PA (MVPA), and overall energy expenditure [14,15], which could present a challenge when providing feedback on the appropriateness of one's behavior. However, this understanding could also be beneficial, as each dimension can be presented as a unique opportunity for behavior change and, in principle, help individuals find bespoke solutions across a person's day, which can help them overcome personal barriers or anchor them to their particular health goals [16]. Moreover, recipients can use reliable multidimensional PA feedback to understand and mitigate against compensatory changes in one aspect of their behavior in response to an attempt to alter another (eg, replacing moderate habitual activity with sedentary time in response to a new exercise regime). Preliminary qualitative data suggest that presenting multiple health-harnessing dimensions to adults is an acceptable, comprehensible, and motivating means of communication that could be readily implemented to support behavior change [17].

The Multidimensional Individualized Physical Activity (MIPACT) trial [18] examined whether a 12-week self-monitoring intervention incorporating multidimensional feedback alongside brief trainer support led to increases in PA behavior among adults at risk of chronic disease. After 3 and 12 months, there was very little change in behavior using this approach despite excellent compliance and adherence [19]. In MIPACT, participants received personal feedback on their multidimensional PA profile and both time spent and energy expended at different PA intensities via the manual upload of data from the monitor to a web-based *app* for viewing their behavior retrospectively. Although this approach is educational and might raise awareness about past behavior [20,21], other persuasive behavioral techniques to support ongoing, acute regulation of behavior or habit formation may be important precursors of sustained change [22,23].

Interventions that have used continuous *real-time* PA feedback (eg, pedometers) have shown promise in supporting changes in PA behavior [24-27]. By extending such work, real-time feedback provided across multiple PA dimensions might compliment a more holistic composite of PA feedback to provide both a *bigger picture* as well as a time-segmented appreciation of PA within the context of people's daily lives. In other words, providing informative data about their progress toward a discrete and achievable activity target in real time can allow people to make quick behavioral adjustments and work toward their overall weekly health goal. The key to such an endeavor is the use of wearable technologies to provide informational feedback and primes based on real-time assessments so as to best capitalize on within-activity motivation *quality* [28].

To best use technological advancements to improve health and well-being, the use of an appropriate motivational theory is a necessity [28]. Self-determination theory (SDT) is a broad and empirically based theory of motivation that provides insight into how to translate informational feedback [29]. At the heart of the SDT is the proposition that people have 3 universal and essential necessities for wellness, healthy functioning, development, and growth, namely the satisfaction of the psychological needs of autonomy, competence, and relatedness [28]. In PA and exercise settings, empirical research has supported the role of need satisfaction in supporting high-quality forms of motivation (ie, autonomous, wherein intrinsic enjoyment and value of the behavior or identified congruence with self-identity guide behavior), better experiences, and higher well-being [28]. Research has also shown autonomous motivation toward exercise to positively predict objectively assessed exercise bouts [30].

Within SDT, it is postulated that when social inputs such as those inherent within interpersonal interactions or embedded in informational, *real-time* feedback satisfy basic psychological needs for autonomy, competence, and relatedness, people are motivated to act for high-quality reasons and experience greater well-being and better experiential outcomes [31]. Applied to the current work, the use of sophisticated PA data visualizations with light touch trainer support, self-monitoring, and real-time feedback was designed to support autonomy (eg, via the provision of choice, exploring new activities or options, and use of meaningful rationales), competence (eg, through the promotion of self-monitoring and clear, constructive, and relevant feedback), and relatedness (eg, demonstrating interest

in people and acknowledging and respecting their perspectives and feelings).

Objective

The primary aim of the present work is to explore whether the provision of sophisticated visual feedback with additional real-time feedback across multiple dimensions of PA supports changes in PA behavior. The secondary aims are to examine whether any changes in behavior lead to meaningful changes in health status over 12 weeks or whether any psychological variables change in response to the intervention. A supplementary aim is to explore the thoughts and feelings of intervention participants to further understand and explain their engagement with and impact of the program.

Methods

Study Design

To explore the efficacy of using combined, multidimensional, composite, and real-time PA feedback on behavior change, a pilot 12-week, 2-armed randomized controlled trial (RCT) design with quantitative and qualitative evaluation was used. The study was registered at ClinicalTrials.gov (NCT02432924) and received ethical approval from the University of Bath's research ethics approval committee for health (reference number: EP 14/15 10). Study outcomes were assessed on 3 occasions. The first 2 assessments were taken before and after a 6-week self-monitoring intervention (or usual behavior if control), with the third assessment following a further 6-week follow-up period in which participants were without feedback. Control participants were offered a 6-week feedback intervention after their third assessment, whereas the intervention group participants were invited to undertake a one-on-one, semistructured interview to provide rich insights into their experience of the intervention.

Participants

Participants were men and women aged between 40 and 70 years who responded to advertisements through the external university webpages, Twitter, and local newspaper articles for people who did not feel they were currently very active. All participants who inquired were sent a participant information sheet and subsequently screened for eligibility via a telephone call. Volunteers were deemed ineligible if they were actively being treated for a chronic disease that might have impeded their ability to change their PA (coronary heart disease, chronic kidney disease [stages 3-5], diabetes mellitus, stroke, heart failure, and peripheral arterial disease) or if they had a PA level (PAL; total energy expenditure divided by resting metabolic

rate) of <2.0, which has been categorized by the World Health Organization as representing a highly active lifestyle [32]. The exploratory nature of this study meant that no formal sample size calculation was undertaken.

Intervention

Waiting List Control Arm

The waiting list control group was encouraged to conduct their usual behavior until they had had 2 further assessments in line with those of the intervention group (ie, 6 and 12 weeks after randomization). At the time of revealing their allocation, waiting list participants were informed that upon completion of the third assessment, they would be able to receive the 6-week self-monitoring intervention in full (without any further follow-up assessment) but to carry on as normal in the meantime.

Intervention Arm (6-Week Active Intervention and 6-Week Follow-up)

Participants randomized to the intervention group returned to the University of Bath at their earliest convenience to undertake a set-up session. Here, participants were shown multidimensional feedback on their weekly PA using the MIPACT web platform, as described by Peacock et al [18]. Briefly, the website provides informational feedback in the form of visual representations of their behavior across a 7-day period. To this end, the feedback encompasses five key health targets (Figure 1A): daily calorie burn, sedentary time, accumulated daily minutes of moderate-intensity activity, weekly MVPA in at least 10-minute bouts, and weekly vigorous-intensity activity accumulated in at least 10-minute bouts. Using a simplified and more detailed graphic, participants were shown each target attainment using a traffic light system where green would indicate a hit target, amber would indicate close to the target, and red would indicate a missed target.

Additional feedback was provided in the form of 24-hour PA patterns that were color coded to indicate the intensity of activity at a given minute of the day (Figure 1B). The web platform also included 2 interactive tabs whereby participants could tag activities to learn about and explore the specific intensity and energy expenditure of a given activity or period (Figure 1C) and forward plan future activities that could be superimposed on a given week's PA patterns to visualize and explore the impact of adding new or existing activities on their health targets (Figure 1D). The Ainsworth Compendium of Physical Activities [33] was used as the basis for calculating the intensity category and personalized energy expenditure for each added activity in the menu of PAs.

Figure 1. Features and examples of feedback and functions included on the Multidimensional Individualized Physical Activity web-based platform. (A) Participants were provided feedback using a traffic light–colored health target attainment schematic across the five dimensions and (B) detailed activity patterns and time use summaries colored in accordance with the intensity of activity during each given minute. (C) Participants were also able to review specific segments of a day to learn about the energy cost and intensity of particular activities and (D) were provided with a planning section where they could see how the addition of new activities, derived from the Ainsworth compendium [33], would affect their health targets if imposed over their existing week.
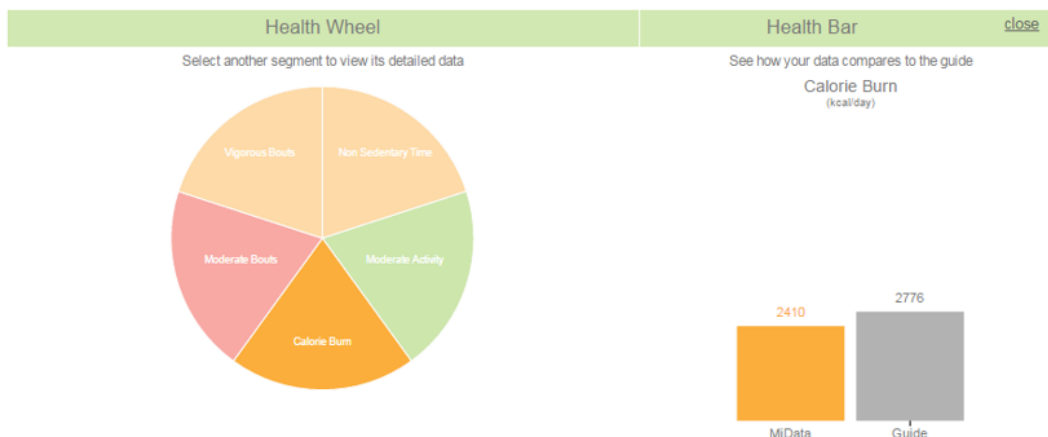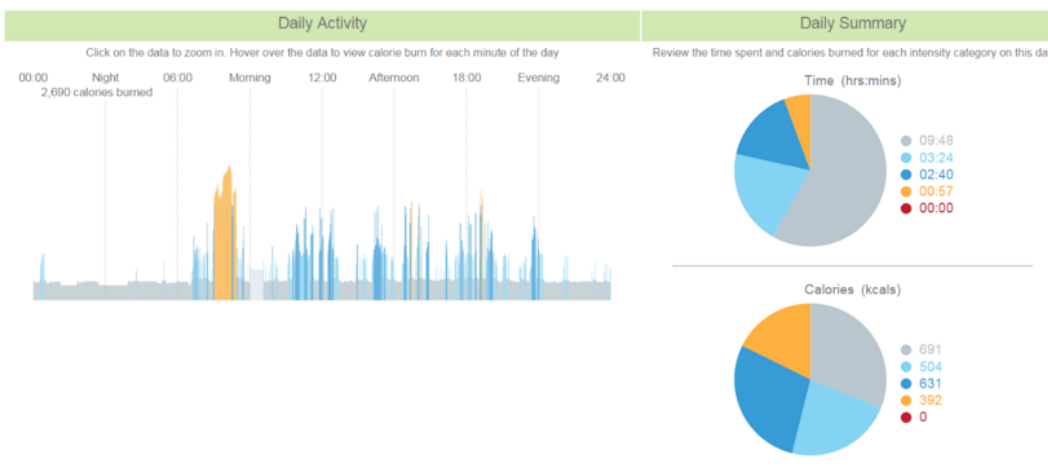
For the real-time feedback element, participants were provided with a Bodymedia Mini (Sensewear; version 8.0) monitor, a smaller model that uses the same algorithms and sensors as the Bodymedia Core used for the assessments, and an accompanying real-time analog display that synced data directly from the armband. The small clip-on display provided feedback on daily accumulated minutes of moderate-intensity activity and minutes of vigorous-intensity activity, calories, and steps that were contextualized alongside the web platform's moderate, vigorous, calorie burn, and nonsedentary time goals, respectively. In

XSL•FO
RenderX

addition to real-time data, the display also stores the total 24-hour values for the previous day and enables users to set personalized targets for each of the 4 activity metrics. If targets were met, a congratulatory message was displayed on the screen, and an alarm sounded to inform the user of their success.

Participants were given an operating manual for the device and encouraged to use it as often as they felt necessary during the 6-week period. Over the course of the intervention period, the participant and researcher met a further 3 times to upload new data from the armband to the MIPACT web platform at weeks 2, 4, and 6. These 15-minute informal sessions afforded each participant the opportunity to troubleshoot any technical queries, get help interpreting their personal multidimensional web feedback, and discuss new plans of action for change. Each session was delivered in a need-supportive manner, encompassing the provision of participant choice; exploration of new activities or plans; and promotion of self-monitoring with clear, constructive, and relevant feedback while taking a clear interest in the perspectives and feelings of the participant [34].

Measurement Procedures

Overview

The baseline laboratory session lasted approximately 45 minutes and involved the signing of informed consent, completion of a questionnaire pack, measurement of brachial seated blood pressure, measurement of anthropometric elements, and retrieval of a fasting venous blood sample from the antecubital vein. Participants were asked to attend the session having abstained from food or caffeine for a minimum of 10 hours. At the end of the session, participants were provided with a PA monitor and instructed to wear the device for 7 consecutive days, removing solely for water-based activities. Participants were also provided with a preaddressed envelope with which to return the activity monitor. The Index of Multiple Deprivation was calculated using participants' postcode on the UK government English indices of deprivation webtool [35], and deprivation decile was extracted for each participant [36]. All procedures other than the signing of informed consent were replicated at the 6- and 12-week follow-up assessment time points.

Primary Outcome: PA

PA was measured using the Bodymedia Mini (Sensewear; version 8.0), which has been shown to accurately measure minute-by-minute energy expenditure [37,38]. To be included in the analysis, participants required a minimum of 6 valid days that included 80% of an assumed 16-hour waking day. On occasions where participants removed the device during sleep or at other times, the estimated resting metabolic rate [39] was assigned to missing data points to complete the 24-hour period. Minute-by-minute energy expenditure was used to determine time (minutes) spent in each of the activity intensity thresholds (sedentary: <1.8 metabolic equivalent of tasks [METs]; light: ≥1.8 and <3.0 METs; moderate: ≥3.0 and <6.0 METs; vigorous: ≥6.0 METs) [40]. These data were used to determine changes in each of the key health-harnessing PA dimensions used in the feedback, including (1) PAL (total energy expenditure divided by resting metabolic rate), (2) sedentary time (percentage of

waking day) and accumulated 1-minute bouts of moderate-intensity activity (minutes per day), (3) MVPA accumulated in bouts of ≥10 minutes (per week), and (5) vigorous-intensity activity accumulated in bouts of ≥10 minutes (per week). The mean daily steps were also determined for each assessment.

Secondary Outcomes: Health Markers

Blood pressure was measured using an automatic sphygmomanometer immediately after 15 minutes of isolated rest. A total of 3 measurements were taken at least 1 minute apart, and the mean of the readings was used as the recorded value. Height was measured without shoes to the nearest millimeter using a Seca stadiometer and weight to the nearest 100 g using a set of digital Tanita (BC-543) scales. These measures were used to calculate BMI (kg/m^2) for each participant. Waist circumference measurements were taken to the nearest millimeter using a Hoechstmass tape measure placed parallel to the floor at the midpoint between the iliac crest and the lowest palpable rib after gentle exhalation. The mean of 3 measurements was taken provided they were within 0.5 cm of one another. A 10 mL fasted venous blood sample was taken at each assessment and used to measure concentrations of plasma glucose, insulin, total cholesterol, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, triglycerides, and C-reactive protein. These metabolic biomarkers were quantified using commercially available spectrophotometric assays (Randox Laboratories, Co) and enzyme-linked immunosorbent assay (serum insulin only: Mercodia AB). The homeostasis model assessment calculator was used to estimate insulin resistance (Homeostatic Model Assessment of Insulin Resistance-2).

Each participant also completed the EuroQol (EQ) 5-dimension 5-level questionnaire [41], which measures the quality of life across five dimensions: mobility, self-care, usual activities, pain or discomfort, and anxiety or depression. The EQ visual analog scale was used to record patients' overall perception of their health from 0 (worst imaginable) to 100 (best imaginable). The Short Form-36 (SF-36) Health Survey Questionnaire was used to determine any changes in perceived physical and mental health [42]. In total, eight health concepts were measured by the SF-36, with four scales each loaded onto two higher-order factors: physical (physical functioning, physical impact on role, bodily pain, and general health) and mental (ie, vitality, social functioning, emotional impact on role, and mental health) health [43]. Using the standardized scoring algorithms outlined by Ware et al [43], component summary scores were computed for physical and mental health ranging from 0 to 100, with higher scores representing better health status.

Secondary Outcomes: Motivation and Psychological Variables

The questionnaire pack included a collection of instruments for which the reliability and validity of the scores have been described at length by the respective cited authors. Where necessary, the stem of the respective questions was altered from its original wording to refer to *PA* rather than *exercise*. To measure participants' motivation, as propagated within SDT, the Psychological Need Satisfaction in Exercise scale [44] was

used to measure autonomy, competence, and relatedness, and the Behavioral Regulation in Exercise Questionnaire-2 [45] was used to explore the participants' motivation to engage in PA (ie, autonomous and controlled reasons). Perceived competence in PA [46] was also included as a more specific measure of an individual's self-belief. The Barrier Self-Efficacy scale [47] was included to determine whether the intervention changed people's confidence to undergo PA in the face of common obstacles, and the Self-Report Habit Index [48] was used to determine the automaticity of PA behavior. The Subjective Vitality Scale [49] was used to detect changes in vitality.

Postintervention Interviews

Participants who successfully completed the intervention were invited to attend a one-on-one semistructured interview to discuss their experience with the program once all follow-up assessments were completed. The topic guide for these interviews (shown in full in Multimedia Appendix 1) included questions to capture participants' views on the utility and retrospective and prospective impact of the intervention for them and unpick the aspects that were most useful and those that might be improved. The interviews typically lasted between 15 and 25 minutes and were recorded using an Olympus digital voice recorder. In addition, all intervention participants completed a feedback form that included rating scales for aspects of the real-time display (overall, personal targets, calories, steps, moderate and vigorous activity) and web-based feedback (overall, health targets, activity patterns, review function, and planning function). Scores ranged from 1=not useful at all and 3=somewhat useful to 5=extremely useful, with a 0 option if the element in question was not used.

Analysis

Mean differences between intervention and control group participants for 6- and 12-week PA and 95% CIs across each of the 6 feedback dimensions were calculated using an analysis of covariance model [50]. Covariates included baseline values of each outcome variable to control for chance imbalances at baseline (accounting for any unequal variance because of unequal group allocation) and the factors used in balancing the groups (sex and weight status) [51]. Bias-corrected and accelerated bootstrapping was used to verify CIs via 5000 replications, as this approach has been recommended to provide more accurate estimates of SEs and CIs with smaller sample sizes [52-54]. The same analysis was used to explore differences

in health outcomes and psychosocial variables at 6 and 12 weeks. Effect sizes (Hedges g) are provided for the mean difference between intervention and control across each variable and are interpreted as 0.2, 0.5, and 0.8 for small, moderate, and large effects, respectively [55]. A post hoc subgroup analysis to explore interactions with covariates observed at 12 weeks was performed, whereby unadjusted means and SDs were calculated to explore whether male versus female and participants with low versus high baseline PA had more pronounced changes in PA data.

Qualitative interviews were interpreted using descriptive deductive and inductive qualitative analyses based on the principles of thematic analysis [56]. Audio files were transcribed verbatim and uploaded to NVivo (version 11; QRS International) for coding and analysis. The lead author, who conducted the interviews, reread through each participant transcript for familiarization and then coded themes within the data. When all transcripts were coded, the themes were compared among participants, and common recurring viewpoints and other important insights were presented in the *Results* section as themes.

Results

Participants

Figure 2 shows the flow of the participants through the study. Of the 102 inquiries, 57 (55.9%) participants were eligible, of whom 5 (9%) were excluded for being too active (PAL ≥2.0) at baseline, and 1 (2%) withdrew because of an allergic reaction to the PA assessment device; therefore, 51 (89%) participants were randomized into either the intervention group or the waiting list control group in a 2:1 allocation ratio to learn more about the intervention. A statistician external to the research team completed randomization and did not disclose any of the details before the completion of recruitment. The statistician stratified the participants by sex (male or female) and weight status (with BMI ≥30 kg/m^2 as the binary cutoff point) using a block size of 6 (which was revealed to the researcher team after the study), giving an overall allocation of 36:15 in favor of the intervention group. No participant withdrew from the study, although one of the intervention participants declined to undergo the end-of-intervention interview. The baseline characteristics of the participants are displayed in Table 1.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram demonstrating participants' progress through the study.
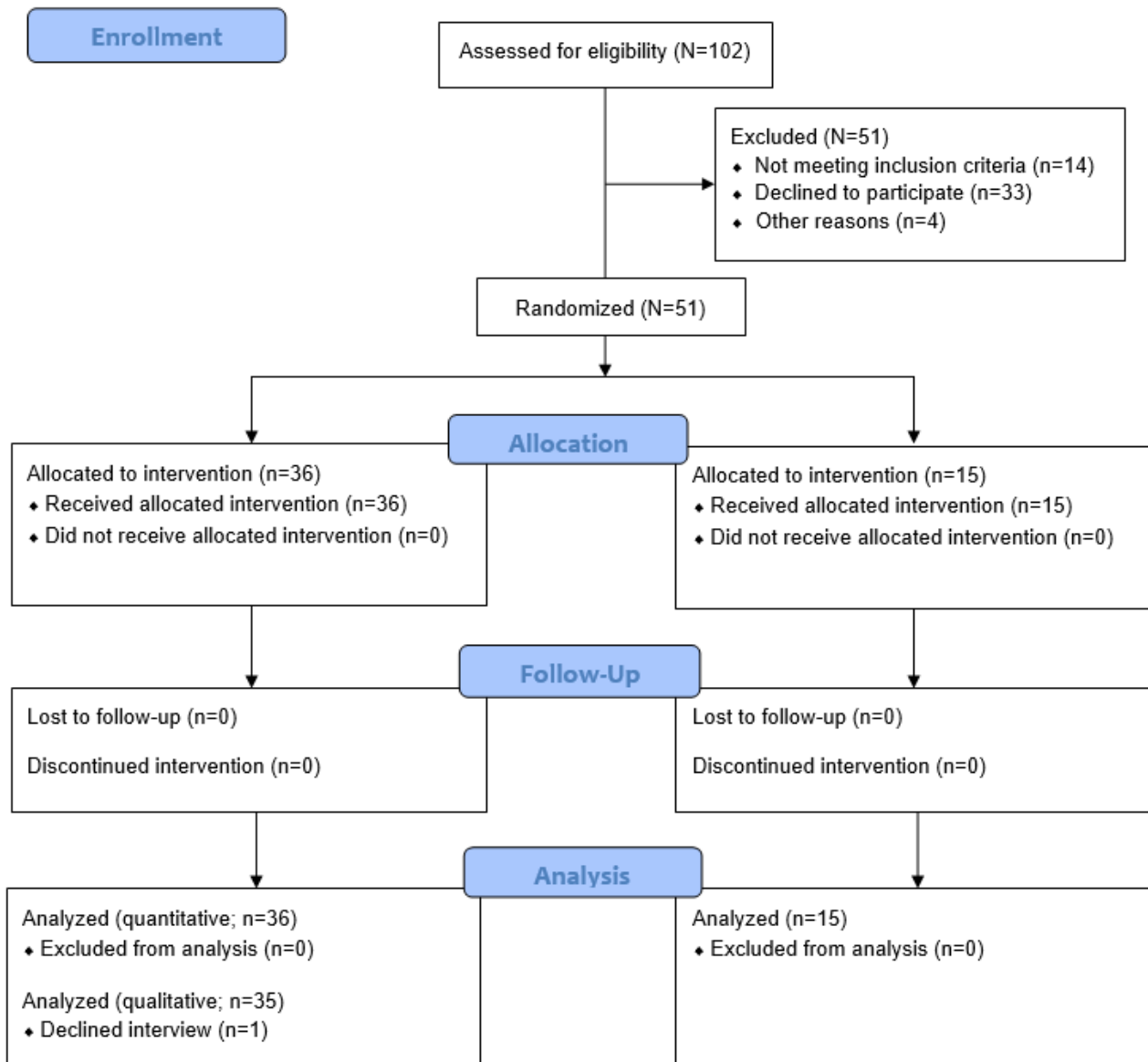
Table 1. Baseline characteristics of study participants (N=51).

Characteristic	All	Intervention (n=36)	Control (n=15)
Age (years), mean (SD)	51.3 (8.4)	52.3 (8.2)	50.1 (8.3)
Age (years), n (%)			
40-55	33 (65)	23 (64)	10 (67)
55-70	18 (35)	13 (36)	5 (33)
Female	28 (55)	20 (55)	8 (53)
Ethnicity (White British)	46 (90)	32 (88)	14 (93)
Marital status, n (%)			
Married or cohabiting	42 (82)	30 (83)	12 (80)
Single, divorced, or widowed	9 (18)	6 (17)	3 (20)
Education, n (%)			
GCSE[a]	3 (6)	3 (8)	0 (0)
A-level	4 (8)	3 (8)	1 (7)
First degree	24 (47)	17 (47)	7 (47)
Higher degree	20 (39)	13 (36)	7 (47)
Index of Multiple Deprivation[b] (decile), mean (SD)	8.0 (2.4)	7.9 (2.3)	8.2 (2.6)
Index of Multiple Deprivation (decile), n (%)			
1-5	11 (22)	8 (23)	3 (20)
6-10	39 (78)	27 (77)	12 (80)
Smoker	2 (4)	0 (0)	2 (13)

[a]GCSE: General Certificate of Secondary Education.

[b]Index of Multiple Deprivation based on postcode calculated [35].

Primary Outcome: PA

All 51 participants provided complete PA data at the 6- and 12-week time points and baseline and were therefore included in the exploratory analysis of the primary outcome. The baseline characteristics of the participants are shown in Table 1. The total 24-hour wear time across the week for the 3 assessment time points was, on average, 98% (SD 1.6%), 96% (SD 8.2%), and 95% (SD 8.1%) in the intervention group and 95% (SD 7.9%), 95% (SD 8.7%), and 94% (SD 12.4%) in the control group, respectively. Table 2 shows the adjusted mean difference (95% CIs) between the intervention and control groups at the 6- and 12-week time points and effect sizes for each PA outcome.

There were no observed differences in any PA outcomes at the 6-week end-of-intervention assessment. At 12 weeks, relative to control participants, the intervention group had reduced mean daily sedentary time by −40 (95% CI −76 to −4) min/day and increased light-intensity activity by 14 (95% CI −78 to 45) minutes per day, moderate-intensity activity by 22 (95% CI 1-45) minutes per day, and vigorous-intensity activity by 2 (95% CI −1 to 6) minutes per day. Post hoc descriptive analysis of subgroups indicated that changes in PA were more pronounced in female participants than in males and for individuals with lower baseline PA levels at 12 weeks (Multimedia Appendix 2, Tables S1 and S2).

Table 2. Mean scores, adjusted mean difference between intervention and control groups, and effect sizes (with 95% CIs) across physical activity dimensions at 6 and 12 weeks.

Outcome and time point	Intervention (n=36)[a], mean (95% CI)	Control (n=15)[a], mean (95% CI)	Adjusted mean difference[a,b] (95% CI)	Effect size, Hedges g (95% CI)
PAL[c,d] (TEE[e] divided by RMR[f])				
Baseline	1.61 (1.55 to 1.66)	1.62 (1.55 to 1.68)	N/A[g]	N/A
Week 6	1.62 (1.57 to 1.67)	1.65 (1.58 to 1.72)	−0.02 (−0.10 to 0.04)	−0.2 (−0.8 to 0.4)
Week 12	1.67 (1.63 to 1.72)	1.58 (1.52 to 1.64)	0.09 (0.02 to 0.15)	0.8 (0.2 to 1.4)
Sedentary time[h] (percentage waking day)				
Baseline	69 (66 to 73)	69 (64 to 73)	N/A	N/A
Week 6	69 (66 to 72)	66 (62 to 70)	3 (−2 to 8)	0.3 (−3 to 0.9)
Week 12	66 (63 to 69)	70 (65 to 74)	−4 (−8 to 1)	−0.5 (−1.1 to 0.1)
Moderate activity[i] (minutes per day)				
Baseline	111 (94 to 129)	117 (99 to 135)	N/A	N/A
Week 6	118 (105 to 130)	127 (107 to 148)	−10 (−28 to 8)	−0.3 (−0.9 to 0.3)
Week 12	132 (118 to 147)	109 (89 to 131)	24 (0 to 45)	0.6 (0.0 to 1.2)
Vigorous bouts[j] (minutes per week)				
Baseline	42 (23 to 65)	26 (12 to 43)	N/A	N/A
Week 6	48 (30 to 70)	46 (24 to 71)	2 (−24 to 28)	0.0 (−0.6 to 0.6)
Week 12	50 (30 to 73)	33 (14 to 55)	18 (−5 to 41)	0.4 (−0.2 to 1.0)
MVPA[k] bouts[j] (minutes per week)				
Baseline	539 (435 to 646)	509 (400 to 622)	N/A	N/A
Week 6	584 (495 to 675)	580 (441 to 725)	4 (−126 to 136)	0.0 (−0.6 to 0.6)
Week 12	658 (571 to 750)	462 (340 to 587)	195 (58 to 331)	0.8 (0.2 to 1.4)
Steps[l] (steps per day)				
Baseline	7403 (6705 to 8093)	7767 (6626 to 8884)	N/A	N/A
Week 6	8207 (7269 to 9114)	8280 (7268 to 9114)	−73 (−1122 to 1017)	0.0 (−0.6 to 0.6)
Week 12	8782 (7987 to 9656)	7236 (6496 to 7991)	1545 (581 to 2553)	0.7 (0.1 to 1.3)

[a]CIs were verified using a bias-corrected and accelerated bootstrap with 5000 replications.

[b]Covariates included stratified randomization factors (BMI at baseline and sex) and baseline scores for the respective outcome variables.

[c]PAL: physical activity level.

[d]Mean total daily energy expenditure divided by daily resting metabolic rate.

[e]TEE: total energy expenditure.

[f]RMR: resting metabolic rate.

[g]N/A: not applicable.

[h]Percentage of waking day.

[i]All minutes ≥3 metabolic equivalents of task.

[j]Activity ≥6 metabolic equivalents of task (vigorous) or ≥3 metabolic equivalents of task (MVPA) accumulated in ≥10 minutes was counted.

[k]MVPA: moderate to vigorous physical activity.

[l]Mean daily step count.

Secondary Outcomes: Health and Well-being

There were no 6- or 12-week differences in any of the cardiometabolic health outcomes measured between the intervention and control groups, except for insulin resistance calculated at week 12. The mental health component summary of the SF-36 improved in the intervention group at 12 weeks; however, neither the SF-S6 nor the physical component summary, EQ 5-dimension 5-level questionnaire, or visual analog scale scores were different at any other time point. The baseline, 6-week, and 12-week scores for all variables are shown in Table 3.

Table 3. Secondary health and psychosocial outcomes at 6 and 12 weeks (N=51)[a].

Outcome and week	Baseline, mean (SD)	Intervention, mean (95% CI[b])	Control, mean (95% CI[b])	Adjusted mean difference (95% CI[b])	Effect size (Hedges *g*)
Health and well-being					
Systolic blood pressure (mm Hg)					
6 weeks	124 (14)	123 (118 to 128)	123 (118 to 128)	0.2 (−4.69 to 4.86)	0.02
12 weeks	124 (14)	124 (120 to 127)	120 (113 to 127)	4.33 (−3.33 to 11.43)	0.43
Diastolic blood pressure (mm Hg)					
6 weeks	86 (10)	86 (83 to 90)	88 (83 to 93)	−1.53 (−6.03 to 3.21)	−0.20
12 weeks	86 (10)	88 (85 to 91)	86 (81 to 91)	1.85 (−3.88 to 7.79)	0.21
Body mass (kg)					
6 weeks	81.9 (14.4)	81.6 (77.8 to 85.4)	82.3 (78.6 to 86.1)	−0.74 (−1.86 to 0.49)	−0.44
12 weeks	81.9 (14.4)	81.8 (78.2 to 85.4)	82.6 (79 to 86.4)	−0.81 (−2 to 0.36)	−0.33
Waist circumference (cm)					
6 weeks	91.8 (11.9)	89.9 (86.6 to 93.3)	90.8 (87.3 to 94.4)	−0.93 (−2.47 to 0.81)	−0.32
12 weeks	91.8 (11.9)	89.2 (86.1 to 92.4)	89.2 (85.7 to 92.9)	0.01 (−1.67 to 1.77)	0.00
Glucose (mmol/L)					
6 weeks	5.3 (0.7)	5.3 (5.1 to 5.5)	5.4 (5.2 to 5.6)	−0.08 (−0.32 to 0.17)	−0.22
12 weeks	5.3 (0.7)	5.3 (5.1,5.5)	5.5 (5.3 to 5.7)	−0.15 (−0.43 to 0.13)	−0.36
Insulin (mIU/mL)					
6 weeks	6.7 (3.8)	6.4 (5.2 to 7.7)	6.6 (5.3 to 8)	−0.18 (−1.85 to 1.7)	−0.06
12 weeks	6.7 (3.8)	6.1 (5.1 to 7.3)	7.3 (6 to 8.9)	−1.25 (−2.4 to −0.16)	−0.50
Insulin resistance					
6 weeks	1.6 (1.1)	1.5 (1.2 to 1.9)	1.6 (1.3 to 2.0)	−0.07 (−0.48 to 0.34)	−0.09
12 weeks	1.6 (1.1)	1.5 (1.2 to 1.7)	1.8 (1.5 to 2.1)	−0.34 (−0.61 to −0.62)	−1.23
Total cholesterol (mmol/L)					
6 weeks	5.6 (0.8)	5.3 (5 to 5.6)	5.5 (5.1 to 5.9)	−0.16 (−0.56 to 0.22)	−0.23
12 weeks	5.6 (0.8)	5.4 (5.2 to 5.6)	5.5 (5.1 to 6)	−0.14 (−0.65 to 0.32)	−0.23
HDL[c] cholesterol (mmol/L)					
6 weeks	1.3 (0.4)	1.4 (1.2 to 1.5)	1.4 (1.3 to 1.5)	−0.04 (−0.17 to 0.11)	−0.18
12 weeks	1.3 (0.4)	1.4 (1.2 to 1.5)	1.4 (1.3 to 1.6)	−0.06 (−0.17 to 0.04)	−0.38
LDL[d] cholesterol (mmol/L)					
6 weeks	3.7 (0.8)	3.5 (3.2 to 3.7)	3.5 (3.2 to 3.8)	−0.07 (−0.38 to 0.25)	−0.11
12 weeks	3.7 (0.8)	3.5 (3.3 to 3.7)	3.5 (3.1 to 3.9)	0.01 (−0.39 to 0.36)	0.03
Triglycerides (mmol/L)					
6 weeks	1.4 (0.8)	1.2 (1 to 1.4)	1.3 (1.1 to 1.6)	−0.12 (−0.39 to 0.14)	−0.27
12 weeks	1.4 (0.8)	1.2 (1.1 to 1.4)	1.5 (1.2 to 1.7)	−0.23 (−0.55 to 0.08)	−0.54
CRP[e] (mg/L)					
6 weeks	2.0 (2.6)	2.4 (1.5 to 3.5)	1.6 (0.9 to 2.2)	0.89 (1.11 to 0.18)	0.32
12 weeks	2.0 (2.6)	1.9 (1.3 to 2.7)	3 (1.6 to 4.5)	−1.09 (−2.8 to 0.51)	−0.48
EQ-5D VAS[f]					
6 weeks	65.2 (16.3)	72.2 (66.8 to 77.3)	71.7 (63.4 to 79.6)	0.45 (−7.57 to 8.45)	0.03
12 weeks	65.2 (16.3)	69.5 (62.5 to 75.7)	71.7 (66.6 to 76.7)	−2.12 (−11.23 to 6.47)	−0.12

XSL•FO

RenderX

Outcome and week	Baseline, mean (SD)	Intervention, mean (95% CI[b])	Control, mean (95% CI[b])	Adjusted mean difference (95% CI[b])	Effect size (Hedges *g*)
EQ-5D-5L[g] score					
6 weeks	0.90 (0.1)	0.92 (0.89 to 0.95)	0.89 (0.85 to 0.93)	0.03 (–0.01 to 0.07)	0.38
12 weeks	0.90 (0.1)	0.89 (0.85 to 0.92)	0.9 (0.85 to 0.95)	–0.01 (–0.06 to 0.04)	–0.16
SF-36[h] physical health					
6 weeks	47.4 (8.4)	51.1 (48.3 to 53.9)	48.5 (45.2 to 51.5)	2.52 (–1.63 to 7.17)	0.39
12 weeks	47.4 (8.4)	47.5 (43.8 to 51.1)	50.1 (46.3 to 53.5)	–2.6 (–7.58 to 2.79)	–0.28
SF-36 mental health					
6 weeks	49.0 (9.8)	50.8 (48.4 to 53)	48.9 (45.3 to 52.6)	1.86 (–2.29 to 6.03)	0.26
12 weeks	49.0 (9.8)	51.7 (47 to 56.3)	43.8 (38 to 49.4)	7.93 (0.74 to 15.18)	0.60
Motivation and psychosocial					
Autonomous motivation					
6 weeks	2.9 (0.7)	3.1 (3 to 3.3)	2.9 (2.7 to 3)	0.26 (0.04 to 0.49)	0.79
12 weeks	2.9 (0.7)	3.1 (2.9 to 3.3)	3 (2.8 to 3.2)	0.1 (–0.1 to 0.3)	0.30
Controlled motivation					
6 weeks	1.5 (0.7)	1.4 (1.2 to 1.6)	1.7 (1.4 to 1.9)	–0.28 (–0.55 to –0.01)	–0.63
12 weeks	1.5 (0.7)	1.3 (1.2 to 1.5)	1.6 (1.3 to 1.8)	–0.2 (–0.5 to 0.09)	–0.42
Overall need satisfaction					
6 weeks	4.7 (1.0)	4.6 (4.2 to 4.9)	4.6 (4.3 to 5)	–0.03 (–0.57 to 0.39)	–0.03
12 weeks	4.7 (1.0)	4.7 (4.5 to 4.9)	4.6 (4.2 to 4.9)	0.12 (–0.21 to 0.43)	0.16
Autonomy					
6 weeks	5.4 (0.6)	5.3 (5.1 to 5.5)	5.4 (5.1 to 5.7)	–0.07 (–0.39 to 0.26)	–0.12
12 weeks	5.4 (0.6)	5.5 (5.3 to 5.7)	5.6 (5.4 to 5.7)	–0.09 (–0.32 to 0.16)	–0.17
Competence					
6 weeks	4.1 (1.2)	4.5 (4.2 to 4.7)	4.1 (3.7 to 4.5)	0.36 (–0.1 to 0.77)	0.47
12 weeks	4.1 (1.2)	4.3 (3.9 to 4.6)	4 (3.5 to 4.4)	0.32 (–0.18 to 0.8)	0.36
Relatedness					
6 weeks	4.3 (1.3)	4.2 (3.7 to 4.6)	4.5 (4.1 to 4.9)	–0.35 (–0.88 to 0.16)	–0.37
12 weeks	4.3 (1.3)	4.4 (3.9 to 4.8)	4.4 (3.9 to 4.8)	0 (–0.74 to 0.72)	0.00
Barrier self-efficacy					
6 weeks	49.7 (16.5)	52.3 (47.6 to 57)	41 (35.1 to 46.8)	11.35 (3.24 to 19.37)	0.84
12 weeks	49.7 (16.5)	53.3 (48 to 58.8)	43.9 (37.3 to 50.3)	9.38 (1.67 to 17.18)	0.68
Vitality					
6 weeks	4.4 (1.1)	5.1 (4.8 to 5.4)	4.3 (3.7 to 4.9)	0.77 (0.17 to 1.37)	0.81
12 weeks	4.4 (1.1)	5.1 (4.7 to 5.5)	4.5 (3.8 to 5.1)	0.57 (–0.04 to 1.2)	0.51
Perceived competence					
6 weeks	5.0 (1.3)	5.3 (4.9 to 5.6)	4.8 (4.4 to 5.1)	0.51 (0.14 to 0.92)	0.70
12 weeks	5.0 (1.3)	5.2 (4.8 to 5.7)	5 (4.6 to 5.5)	0.2 (–0.41 to 0.9)	0.19
Habit					
6 weeks	1.6 (1.0)	2.1 (1.8 to 2.4)	1.5 (1.2 to 1.9)	0.56 (0.16 to 0.97)	0.84

Outcome and week	Baseline, mean (SD)	Intervention, mean (95% CI[b])	Control, mean (95% CI[b])	Adjusted mean difference (95% CI[b])	Effect size (Hedges g)
12 weeks	1.6 (1.0)	2.1 (1.8 to 2.3)	1.7 (1.3 to 2)	0.41 (−0.04 to 0.86)	0.58

[a]Covariates include baseline score for each parameter (as indicated in the pooled mean baseline column), BMI, and sex.

[b]CIs verified using bias-corrected and accelerated bootstrapping with 5000 repetitions.

[c]HDL: high-density lipoprotein.

[d]LDL: low-density lipoprotein.

[e]CRP: C-reactive protein.

[f]EQ-5D VAS: EuroQol 5-dimension visual analog scale.

[g]EQ-5D-5L: EuroQol 5-dimension 5-level questionnaire.

[h]SF-36: Short Form-36.

Secondary Outcomes: Motivation and Psychosocial Variables

Relative to the control group, the intervention group had a reduction in controlled behavioral regulation (external and introjected regulation), and increases in autonomous behavioral regulation (intrinsic, integrated and identified regulation), perceived competence for PA and habit strength at the 6-week assessment but not at 12 weeks. Barrier self-efficacy was increased in the intervention group at 6 weeks and was sustained at the 12-week follow-up. Subjective vitality was also increased in the intervention group at 6 weeks but was not sustained until 12 weeks. No changes in overall psychological need satisfaction or its subscales were observed (Table 3).

Intervention Component Evaluation

Participants were asked to provide their subjective ratings of the *usefulness* of intervention features at the 6-week assessment and qualitative feedback following their 12-week assessment. From the subjective ratings participants ranked the real-time display (mean 4.5, SD 0.8) higher than the web-based MIPACT platform (mean 3.3, SD 1.5) using a scale from 1 *not useful at all* to 5 *very useful*, or 0=*not used*. Each aspect of the real-time display consistently rated as more useful than the features of the web platform. Specifically, the display of calorie data (mean 4.0, SD 0.9), steps (mean 4.3, SD 0.7), moderate-vigorous activity (mean 4.3, SD 1.2) and having personal targets (mean 4.0, SD 1.5) was rated higher than the composite health target

(mean 3.6, SD 1.5) and activity pattern (mean 3.7, SD 1.5) data. The lowest-ranked features were the more interactive web-based tools, namely the review (mean 2.3, SD 2.0) and planning (mean 2.0, SD 1.9) sections of the MIPACT website.

Qualitative Evaluation

Qualitative feedback offers further insight into these ratings. Textbox 1 provides a summary of the key themes identified in the analysis of intervention participants' interviews and quotations to illustrate each theme. All intervention participants championed the feedback as useful for raising their consciousness and awareness of their own PA behavior, with many mentioning an improved understanding of the time they spent inactive (theme 1). More than half of the intervention participants postulated that PA was now more of a priority after having been through the program and that it reinforced their belief that PA was a means of improving health (theme 2). The self-monitoring element helped individuals gauge how much PA was required to meet certain health recommendations (theme 3). According to many of the participants, the program inspired them to increase their PA levels, and two-thirds alluded to the fact that the multidimensional nature of the feedback assisted them in finding personal solutions (theme 4). Some participants said that during the 6-week program, they would consciously go out of their way to achieve the targets, and many put added emphasis on steps as a key and achievable daily motivator (theme 5).

Textbox 1. Key themes identified in the qualitative analysis and example quotations (participant information provided as sex, age, baseline physical activity level).

Theme 1: personalized feedback improved understanding of one's own behavior

- *I think it's, it's changed, it's changed my day-to-day activity, and I am a lot more conscious of the fact that I am sitting a lot, and part of it, there was a realisation that I wasn't very active.* [male, 46 years, 1.48]

- *Yeah, I think I was probably overestimating what I was doing, I thought I was more active than I was in a way so...when you see it's like oh you are actually doing as much as I thought I'd probably on my feet but I'm not necessarily so doing anything that is going to benefit me stop so yeah it's definitely made me more aware of the need.* [female, 42 years, 1.42]

Theme 2: physical activity is now more of a priority or reinforced importance

- *Um, well it certainly hasn't become any less important. I probably would say that it has become more important because the awareness breeds that sort of feeling, you know, that this is something that is not just a one-off, you know. Over a three-month period, it's, it's life and it should continue.* [male, 59 years, 1.72]

- *And then hopefully, my hope is, as ie, as I lose weight...Because that's one thing I haven't done is lost weight...um, is once I have lost more weight that I will feel fitter and then I can up that target. But I don't want to try and do too much, too soon.* [female, 62 years, 1.25]

Theme 3: feedback helped people understand how to meet recommendations

- *I found it interesting, you know? Because I know how many steps it takes me to go down our town and round and back to the house it's at about 1800, I think. And I know how many is to go to the railway and things like that.* [female, 63 years, 1.37]

- *But of course, that whole thing then tipped me nicely over and I was...So, it had that useful upturn, and equally, as I said before, it helps me gauge just exactly how much distance I need to be covering to meet a, sort of a standard target.* [male, 48 years, 1.65]

Theme 4: feedback helped motivate and find personalized solutions to increase physical activity

- *So, I knew that I had to just get back into doing something...And having that monitor was almost like a critical friend, it was there to say 'you can do this.'* [female, 48 years, 1.55]

- *Yes, that really helped and then over the six-week period, every week I was trying to do a little bit more and like I say, it's not very difficult to do it's just that now you are conscious of it and you are aware of it that you have to achieve so many steps per day.* [male, 48 years, 1.87]

- *Definitely. Anything is worth it. Any...Any activity, it doesn't have to be gym five days a week. If I wasn't doing five days a week at the gym, I didn't consider myself to be active, basically. So now I know that because I wasn't training five days a week, and I was actually able to show some green lights when I wasn't doing the fi...it makes me realise that all of it counts. It's completely changed. I'm actually more active because I'm down on myself for not doing five days a week at the gym.* [female, 44 years, 1.52]

- *And it's achievable without knocking myself flat you know I can do it in little steps and I can move myself forward in little ways rather than try and charge at a wall and break through a wall. It is much easier that way. So again, using the word empowerment it has sort of empowered me into thinking I could do this.* [male, 54 years, 1.39]

Theme 5: real-time feedback prompted attainment of acute daily goals

- *Um, and I did find it motivating, and I did, um, you know, I was known to leave the house at kind of five minutes to bedtime to walk around the block at the time...Or spend five minutes doing star jumps to try and get vigorous activity in. So yes, having the targets I found very helpful, and yeah, and motivating and interesting and fun.* [female, 56 years, 1.56]

- *Um, I did actually, I surprised myself in how easy it was to make step goals. I did not think I walked that much but as soon as I was just tracking it, it was like "actually I'm not far off daily amounts if I just do a little bit more and better hit that target.* [male, 41 years, 1.53]

Theme 6: now able to fit more physical activity into routine

- *Making a conscious almost, not a plan, plan is probably a bit too grand, but saying 'right each week I must do a certain amount of activity' and I plan that and think about it and so...The type of person I am, I'm quite a sort of structured and quite organised person so just building that into my routine is a change in my behaviour.* [male, 53 years, 1.86]

Theme 7: injury and illness hampered progress during intervention

- *Right, um, it was slightly complicated by the fact that I was ill right in the middle of it so...I started off really motivated and felt really good about it and it was building very well. And then unfortunately, after about a month I guess, I got this fluey type thing, which really did kick in and, made it a bit of a struggle to do as much and build as much as I wanted to do it. And then of course it's sort of came to the end of the programme really so I don't feel like I did it as much justice as I would have liked to have done.* [female, 67 years, 1.58]

Theme 8: felt confident in keeping up or increase physical activity further

- *Yeah, it definitely made me think a bit more about the moderate bouts of activity and how important they are. And it made me more keen to do things like walking the dogs and, you know, walks to school and I wouldn't the thought that to be useful before. And I think 'oh they are quite a useful way of getting in extra steps.'* [female, 43 years, 1.71]

Theme 9: intervention led to improved confidence and enjoyment of physical activity

- *Yeah. I mean, variety...yeah, I think that's been really helpful, actually. Because it's less boring and, um, you don't perceive it as...I think my perception of what exercise was and what it actually is very different now. So now activity isn't exercise. Activity is just anything.* [female, 45 years, 1.52]

- *Knowing how more confident I am, which I, perhaps if I wasn't recording it somewhere I wouldn't have been aware of that...So that's erm, yeah that's a nice position to be in, having seen confidence increase with various things, various types of activities, it's a nice position to be in for sort of in the future.* [female, 48 years, 1.51]

Theme 10: intervention prompted participants to purchase or consider purchasing a real-time feedback device

- *It has spurred me on to get one of these, to actually buy one of these Fitbits. Which is just going to continue to let me know in real-time exactly what I'm doing and very similar in fact it is in steps and calories burnt off and what have you and the fact everyone else in the office have got one.* [female, 60 years, 1.52]

- *Which I suppose sounds really obvious when you say it, but it hadn't ever linked with me before. And, I now have a little Fitbit because I want to now...I've become slightly obsessed with steps.* [female, 63 years, 1.37]

Theme 11: real-time feedback element considered the most useful element of the intervention

- *The instant display I think is what...I mean, I did go online that that's in retrospect, you didn't get to see that until you had already done it. Whereas in today's society we want instant answers so having the display and being able to look at it, um, was, you know, was motivating.* [female, 62 years, 1.25]

- *Um, the, the monitoring device, I found I used the little tiny daily, daily monitor, all the time...that was almost obsessive!* [male, 46 years, 1.48]

Theme 12: web-based feedback useful initial picture

- *You see on the computer screen and it was just flat line, I think that that is, visually, or when you look at it and you look at the figures and you look at that...that had probably quite an impact. And I think that that is...Probably the wake-up call which will remain with me, yeah, visually seeing it.* [male, 64 years, 1.72]

Theme 13: web-based component could be improved

- *Because I could only look at it at certain times at home without being able to do it when I wanted to do it was frustrating...If you see what I mean? So, just only having a sort of biweekly uploaded my information...I wish I could have just done it as and when and seen more feedback.* [male, 41 years, 1.53]

- *I think to be perfectly honest it was...it was sort of time element of it. I didn't feel that I had the time just to sit and...and look at it, which I probably should have done, but it felt as if the more instantaneous response from the monitor was actually...or the display was...was what I needed on a day-to-day basis.* [female, 52 years, 1.64]

Theme 14: some issues with device comfort or data trust

- *The exercise I tend to do is like cycling and the bottom half of my body it probably won't show a great deal of vigorous activity. Which, okay it is the limitation of the technology and the technology at that time, but I was mildly irritated by that.* [male, 55 years, 1.50]

Theme 15: suggested improvements

- *10,000 steps is nice and easy cause that's just walking, you can just incorporate that into your daily activity, but then the vigorous activity, I could do it if I go for a run, but any other way I wouldn't know. I only had ideas of cycling and rowing and though there are suggestions, but a programme of how you can achieve them would have been helpful.* [female, 45 years, 1.60]

- *So, if you said to me your cholesterol is 5 at the end of the study you told me my cholesterol...well...I found out my cholesterol...because you know cholesterol response to exercise had dropped to 3.5 that would have been a big encouragement.* [female, 55 years, 1.61]

Most of the intervention group felt that they were now being more proactive about fitting PA into their routine after the intervention (theme 6), whereas a handful of participants alleged to have had an illness or injury during the program that hampered their progress (theme 7). Approximately two-thirds of participants expressed further intentions to take up new, or perform more, PA, and approximately half of the group felt confident that they could maintain their PA levels after the program and had made lasting behavioral changes (theme 8). In addition, some participants felt that they had improved their confidence and sense of competence for PA, whereas others expressed a greater enjoyment for PA and an improved sense of health and well-being (theme 9). Many participants said that they missed not having the real-time activity monitor once it was removed after 6 weeks, and by the time of the interview, approximately one-third of participants had purchased a commercial PA tracker for personal use, with many more considering acquiring a device (theme 10).

For many of the intervention participants, the real-time display was a favorable component for the self-monitoring of activity and more important than website feedback (theme 11). That said, there was still a reasonable proportion of participants who made reference to the multidimensional feedback as a useful way of viewing the overall picture, and some even described it as a wake-up call (theme 12). Some participants suggested that

their engagement with the feedback on the web platform may have been improved if it was more readily available and that sitting down at a computer felt counterintuitive to being (more) physically active (theme 13). The data also revealed that for certain participants, there were minor issues with the device itself in terms of either trusting the feedback or its wearability (theme 14). Finally, a handful of participants made recommendations for the improved utility of the monitor and feedback system, which included the need for more prompts and guidance or links to their health data collected during assessment sessions to help them evaluate the impact of more PA or for motivation (theme 15).

Discussion

Principal Findings

In this exploratory RCT, we evaluated a 6-week intervention using personalized real-time digital PA feedback and sophisticated web-based multidimensional PA feedback combined with brief trainer support. Exploratory analysis demonstrated no change in PA between groups immediately after the intervention; however, improvements were found for several PA metrics that formed part of the feedback at the 12 weeks follow-up. Subgroup analysis suggests that this effect was more pronounced in female participants and in those with lower baseline activity levels. Very little changed in respect to secondary health outcomes, with the exception of insulin resistance and self-reported mental health, which showed signs of improvement after 12 weeks. Qualitative data suggest that participants found the multicomponent intervention informative and motivating, with the real-time feedback being heralded as the single most memorable and supportive component within the context of the overall treatment package.

Comparison With Other Literature

A novel aspect of this study was the multidimensional approach that, we hypothesized, helps individuals to understand their behavior and find bespoke behavioral solutions for increasing their PA [16]. We hypothesized that using a multidimensional approach to PA promotion and feedback would provide options and self-endorsed choices to foster autonomous motivation and would satisfy the needs for autonomy and competence. Following the 6-week active phase of the intervention, we witnessed favorable improvements in autonomous versus controlled motivation, perceived competence, and barrier self-efficacy, which offers support for the proposed mechanism and the multidimensional approach. Moreover, the qualitative evaluation aligns with our previous development work, which found that receiving detailed, visual multidimensional PA feedback is helpful for raising awareness, understanding, and intention to change [17,20]. We hypothesized that the addition of real-time feedback might help translate those intentions into behavior [57,58]. However, beneficial differences in PA were only observed after a 6-week period in which the whole treatment package (including the real-time display) was removed rather than immediately after the active intervention. Participants expressed that they valued the real-time feedback during the interview more than other components of the intervention and highlighted that it empowered them to adjust their behavior on

a more discrete basis as they strive toward a desired daily target (eg, serving as a prompt to take an additional 1000-step walk if they were short toward the end of a day).

Other studies have observed real-time feedback to be an effective tool for increasing PA when used in conjunction with detailed web-based feedback and trainer support. Vandelenotte et al [27] demonstrated that adding a Fitbit device to their theory-informed web-based PA intervention increased total PA and MVPA by up to 3 times relative to a nontracker, web-only group. Their study, whose website went beyond the provision of feedback to provide individually tailored advice on a number of self-regulation strategies, found that real-time monitoring also improved engagement and adherence to the main web content and the overall package of behavioral support. Similarly, a large RCT by Harris et al [25] found that combining brief nurse support, retrospective accelerometer feedback, and continuous pedometer feedback led to significant, sustained changes in PA in the intervention versus control groups at 3 and 12 months. In another trial, the same research team demonstrated that continuous pedometer feedback provided effective support both with and without trainer input versus a control group with no feedback or trainer support [24]. The effects observed in these studies, albeit modest in size, were maintained after 3- to 4-year follow-up periods [59], suggesting that technology-based PA interventions such as the one used in this study can help individuals make long-lasting changes.

Our qualitative findings corroborated key findings from the Pedometer And Consultation Evaluation-U and Pedometer Accelerometer Consultation Evaluation-Lift studies, which were conducted by Harris et al [24,25]. Specifically, participants who received the nurse-led pedometer intervention experienced greater awareness of the PA guidelines and their own PA levels. They also placed more importance on being active and helped participants to embed PA in their own routines [25,60,61]. Participants also found real-time feedback useful for initiating and monitoring behavior change in relation to personalized goals, and, mirroring the findings reported in the present work, some went on to invest in other wearable trackers after their intervention, although distrust in the accuracy was identified as a potential barrier to effectiveness [60]. A set of themes derived from this study (eg, illness and injury) and the work of Harris et al [25] (eg, weather and lack of time) was the fact that common external barriers still existed for participants that could not be overcome by the real-time feedback interventions. Recommendations from participants in these and other qualitative studies suggest that more interpersonal prompts and guidance, resources for planning activities, meaningful challenges, and links to health data may be avenues to overcome barriers and enhance intrinsic motivation and behavioral maintenance in real-time feedback-based interventions [62-64].

The incongruent findings observed at 6 (immediately after the intervention, no difference) and 12 weeks (after a 6-week follow-up, moderate to large effects) warrant further consideration. The assessment used in this study and most RCTs with device-based PA outcomes relied on weekly snapshots of participants' behavior. The small sample size and variability around the mean scores of the control group suggest that any fluctuations might be because of noise in the assessment. In the

intervention group, a weekly snapshot may not give an accurate representation of a person's true behavior [65]. Continuous measurement in both intervention and control groups would help decipher whether the 6-week observation is, for example, indicative of a dip in behavior following the removal of feedback, or whether the 12-week observations is, for example, indicative of the intervention participants learned rather than new habitual behavior. Given the advancing technology that enables long-term data capture, future studies would do well to investigate the stability and representativeness of PA behavior to guide trials on the most appropriate assessment window.

Strengths and Limitations

The strengths of this study include the almost complete 24-7 objective, PA assessment and high compliance to the intervention, measured as the completeness of attendance to upload sessions and PA monitor wear time in the intervention group, and assessments in all groups (all 100%). The use of quantitative and qualitative evaluations also provides rich insights into the effectiveness of this approach. Limitations include the small sample size, short follow-up period, and use of a nonclinical population, which prevents the performance of more robust statistical analyses and means that any interpretation of these results should be viewed as preliminary rather than definitive and generalizable.

There is also a need to improve the synchronicity of the wearable devices as, in this study, technical issues meant that global web-based feedback could not be fully self-monitored without the trainer needing to recalibrate the personal targets and user profile used within the real-time display. This lack of autonomy over the web platform use may contribute to a more favorable evaluation of the real-time feedback element. Accordingly, we can determine neither the respective contributions of the real-time, web-based, or trainer support on individual participants' behavior change nor whether favorable perceptions of the real-time element would have been the same without the more comprehensive web-based feedback. Recent meta-analyses of SDT-based techniques support the notion that different self-regulatory and trainer-delivered strategies may be useful for optimizing an individual's motivation for PA [66,67]. Therefore, it is unlikely that any single component will be effective in isolation and that multicomponent interventions are required to provide optimal behavioral support. Nonetheless, future trials using more adaptable, multiple-group designs such as the multiphase optimization strategy would be advised to augment the complex intervention and evaluate the relative and complementary importance of the different elements [68].

Conclusions

In conclusion, this exploratory RCT represents the first attempt at combining multidimensional feedback with real-time data and light touch trainer support across several important health-harnessing dimensions of PA as a means of helping individuals change their behavior. The results suggest that this approach may be a useful strategy for helping individuals with low levels of PA change their behavior. These findings can inform the design of future studies with larger and more diverse sample sizes, detailed process evaluations, and longer follow-up periods to explore the effectiveness of real-time, multidimensional feedback.

Conflicts of Interest

None declared.

Multimedia Appendix 1
End-of-study semistructured interview topic guide and evaluation form.
[DOCX File , 118 KB - formative_v6i3e26525_app1.docx]

Multimedia Appendix 2
Subgroup data for physical activity outcomes.
[DOCX File , 26 KB - formative_v6i3e26525_app2.docx]

References

1. Kyu HH, Bachman VF, Alexander LT, Mumford JE, Afshin A, Estep K, et al. Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. Br Med J 2016 Aug 09;354:i3857 [FREE Full text] [doi: 10.1136/bmj.i3857] [Medline: 27510511]

2.   Lear S, Hu W, Rangarajan S, Gasevic D, Leong D, Iqbal R, et al. The effect of physical activity on mortality and cardiovascular disease in 130 000 people from 17 high-income, middle-income, and low-income countries: the PURE study. Lancet 2017 Dec;390(10113):2643-2654. [doi: 10.1016/S0140-6736(17)31634-3]

3.   Teychenne M, Ball K, Salmon J. Physical activity and likelihood of depression in adults: a review. Prev Med 2008 May;46(5):397-411. [doi: 10.1016/j.ypmed.2008.01.009] [Medline: 18289655]

4.   Owen N, Healy GN, Dempsey PC, Salmon J, Timperio A, Clark BK, et al. Sedentary behavior and public health: integrating the evidence and identifying potential solutions. Annu Rev Public Health 2020 Apr 02;41(1):265-287. [doi: 10.1146/annurev-publhealth-040119-094201] [Medline: 31913771]

5.   Tremblay MS, Colley RC, Saunders TJ, Healy GN, Owen N. Physiological and health implications of a sedentary lifestyle. Appl Physiol Nutr Metab 2010 Dec;35(6):725-740. [doi: 10.1139/h10-079]

6.   Cheney S. Relationship of sedentary behavior and physical activity to incident cardiovascular disease: results from the women's health initiative. J Emerg Med 2013 Sep;45(3):481. [doi: 10.1016/j.jemermed.2013.07.008]

7.   Ding D, Lawson KD, Kolbe-Alexander TL, Finkelstein EA, Katzmarzyk PT, van Mechelen W, et al. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. Lancet 2016 Sep;388(10051):1311-1324. [doi: 10.1016/s0140-6736(16)30383-x]

8.   Chiauzzi E, Rodarte C, DasMahapatra P. Patient-centered activity monitoring in the self-management of chronic health conditions. BMC Med 2015 Apr 09;13(1):77 [FREE Full text] [doi: 10.1186/s12916-015-0319-2] [Medline: 25889598]

9.   Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016 Feb 2;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]

10. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. J Am Med Assoc 2015 Feb 03;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]

11. Standage M, Curran T, Rouse PC. Self-determination- based theories of sport, exercise, and physical activity motivation. In: Horn TS, Smith AL, editors. Adv Sport Exerc Psychol. London, UK: Bloomsbury Publishing; 2019:289-311.

12. Gabriel KK, Morrow JR, Woolsey AT. Framework for physical activity as a complex and multidimensional behavior. J Phys Act Health 2012 Jan;9 Suppl 1:S11-S18. [doi: 10.1123/jpah.9.s1.s11] [Medline: 22287443]

13. Thompson D, Batterham AM. Towards integrated physical activity profiling. PLoS One 2013 Feb 20;8(2):e56427 [FREE Full text] [doi: 10.1371/journal.pone.0056427] [Medline: 23437131]

14. Thompson D, Batterham AM, Markovitch D, Dixon NC, Lund AJ, Walhin J. Confusion and conflict in assessing the physical activity status of middle-aged men. PLoS One 2009 Feb 2;4(2):e4337 [FREE Full text] [doi: 10.1371/journal.pone.0004337] [Medline: 19183812]

15. Scheers T, Philippaerts R, Lefevre J. Compliance with different physical activity recommendations and its association with socio-demographic characteristics using an objective measure. BMC Public Health 2013 Feb 14;13(1):136 [FREE Full text] [doi: 10.1186/1471-2458-13-136] [Medline: 23409982]

16. Thompson D, Peacock O, Western M, Batterham AM. Multidimensional physical activity: an opportunity, not a problem. Exerc Sport Sci Rev 2015 Apr;43(2):67-74 [FREE Full text] [doi: 10.1249/JES.0000000000000039] [Medline: 25607280]

17. Western MJ, Peacock OJ, Stathi A, Thompson D. The understanding and interpretation of innovative technology-enabled multidimensional physical activity feedback in patients at risk of future chronic disease. PLoS One 2015 May 4;10(5):e0126156 [FREE Full text] [doi: 10.1371/journal.pone.0126156] [Medline: 25938455]

18. Peacock OJ, Western MJ, Batterham AM, Stathi A, Standage M, Tapp A, et al. Multidimensional individualised Physical ACTivity (Mi-PACT)--a technology-enabled intervention to promote physical activity in primary care: study protocol for a randomised controlled trial. Trials 2015 Aug 28;16(1):381 [FREE Full text] [doi: 10.1186/s13063-015-0892-x] [Medline: 26314577]

19. Peacock OJ, Western MJ, Batterham AM, Chowdhury EA, Stathi A, Standage M, et al. Effect of novel technology-enabled multidimensional physical activity feedback in primary care patients at risk of chronic disease - the MIPACT study: a randomised controlled trial. Int J Behav Nutr Phys Act 2020 Aug 08;17(1):99 [FREE Full text] [doi: 10.1186/s12966-020-00998-5] [Medline: 32771018]

20. Western MJ, Thompson D, Peacock OJ, Stathi A. The impact of multidimensional physical activity feedback on healthcare practitioners and patients. BJGP Open 2019 Feb 05;3(1):bjgpopen18X101628. [doi: 10.3399/bjgpopen18x101628]

21. Godino JG, Watkinson C, Corder K, Marteau TM, Sutton S, Sharp SJ, et al. Impact of personalised feedback about physical activity on change in objectively measured physical activity (the FAB study): a randomised controlled trial. PLoS One 2013 Sep 16;8(9):e75398 [FREE Full text] [doi: 10.1371/journal.pone.0075398] [Medline: 24066178]

22. Rhodes RE, Yao CA. Models accounting for intention-behavior discordance in the physical activity domain: a user's guide, content overview, and review of current evidence. Int J Behav Nutr Phys Act 2015 Feb 07;12(1):9 [FREE Full text] [doi: 10.1186/s12966-015-0168-6] [Medline: 25890238]

23. French DP, Cameron E, Benton JS, Deaton C, Harvie M. Can communicating personalised disease risk promote healthy behaviour change? A systematic review of systematic reviews. Ann Behav Med 2017 Oct 13;51(5):718-729 [FREE Full text] [doi: 10.1007/s12160-017-9895-z] [Medline: 28290066]

24. Harris T, Kerry SM, Limb ES, Victor CR, Iliffe S, Ussher M, et al. Effect of a primary care walking intervention with and without nurse support on physical activity levels in 45- to 75-year-olds: the pedometer and consultation evaluation (pace-up)

cluster randomised clinical trial. PLoS Med 2017 Jan 3;14(1):e1002210 [FREE Full text] [doi: 10.1371/journal.pmed.1002210] [Medline: 28045890]

25.   Harris T, Kerry SM, Victor CR, Ekelund U, Woodcock A, Iliffe S, et al. A primary care nurse-delivered walking intervention in older adults: PACE (pedometer accelerometer consultation evaluation)-Lift cluster randomised controlled trial. PLoS Med 2015 Feb 17;12(2):e1001783 [FREE Full text] [doi: 10.1371/journal.pmed.1001783] [Medline: 25689364]

26.   Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. J Am Med Assoc 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]

27.   Vandelanotte C, Duncan MJ, Maher CA, Schoeppe S, Rebar AL, Power DA, et al. The effectiveness of a web-based computer-tailored physical activity intervention using fitbit activity trackers: randomized trial. J Med Internet Res 2018 Dec 18;20(12):e11321 [FREE Full text] [doi: 10.2196/11321] [Medline: 30563808]

28.   Standage M, Ryan RM. Self-determination theory in sport and exercise. In: Handbook of Sport Psychology (4th Ed). Hoboken, New Jersey, United States: John Wiley & Sons; 2019:352-378.

29.   Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037/0003-066x.55.1.68]

30.   Teixeira PJ, Carraça EV, Markland D, Silva MN, Ryan RM. Exercise, physical activity, and self-determination theory: a systematic review. Int J Behav Nutr Phys Act 2012 Jun 22;9(1):78 [FREE Full text] [doi: 10.1186/1479-5868-9-78] [Medline: 22726453]

31.   Standage M, Sebire SJ, Loney T. Does exercise motivation predict engagement in objectively assessed bouts of moderate-intensity exercise? A self-determination theory perspective. J Sport Exerc Psychol 2008 Aug;30(4):337-352. [doi: 10.1123/jsep.30.4.337] [Medline: 18723896]

32.   Livesey G. Energy and protein requirements the 1985 report of the 1981 Joint FAO/WHO/UNU Expert Consultation. Nutr Bulletin 1987 Sep;12(3):138-149. [doi: 10.1111/j.1467-3010.1987.tb00040.x]

33.   Ainsworth B, Haskell WL, Herrmann SD, Meckes N, Bassett DR, Tudor-Locke C, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. Med Sci Sports Exerc 2011 Aug;43(8):1575-1581. [doi: 10.1249/MSS.0b013e31821ece12] [Medline: 21681120]

34.   Teixeira PJ, Marques MM, Silva MN, Brunet J, Duda JL, Haerens L, et al. A classification of motivation and behavior change techniques used in self-determination theory-based interventions in health contexts. Motivat Sci 2020 Dec;6(4):438-455. [doi: 10.1037/mot0000172]

35.   English indices of deprivation 2015. Ministry of Housing Communities & Local Government. 2015. URL: https://imd-by-postcode.opendatacommunities.org/imd/2015 [accessed 2022-01-18]

36.   Smith T, Noble M, Noble S, Wright G, McLennan D, Plunkett E. The English Indices of Deprivation. London, UK: Department for Communities and Local Government; 2015.

37.   Lee J, Kim Y, Welk GJ. Validity of consumer-based physical activity monitors. Med Sci Sports Exerc 2014 Sep;46(9):1840-1848. [doi: 10.1249/MSS.0000000000000287] [Medline: 24777201]

38.   Johannsen D, Calabro MA, Stewart J, Franke W, Rood JC, Welk GJ. Accuracy of armband monitors for measuring daily energy expenditure in healthy adults. Med Sci Sports Exerc 2010 Nov;42(11):2134-2140. [doi: 10.1249/MSS.0b013e3181e0b3ff] [Medline: 20386334]

39.   Schofield WN. Predicting basal metabolic rate, new standards and review of previous work. Hum Nutr Clin Nutr 1985;39 Suppl 1:5-41. [Medline: 4044297]

40.   Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, American College of Sports Medicine, American Heart Association. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Circulation 2007 Aug 28;116(9):1081-1093. [doi: 10.1161/CIRCULATIONAHA.107.185649] [Medline: 17671237]

41.   Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001 Jul 08;33(5):337-343. [doi: 10.3109/07853890109002087] [Medline: 11491192]

42.   Ware JE, Sherbourne CD. The MOS 36-ltem Short-Form Health Survey (SF-36). Med Care 1992;30(6):473-483. [doi: 10.1097/00005650-199206000-00002]

43.   Ware Jr JE, Kosinski M, Keller S. SF-36 physical and mental health summary scales: a user's manual. New England Medical Center. 1994. URL: https://www.researchgate.net/profile/John-Ware-6/publication/292390260_SF-36_Physical_and_Mental_Health_Summary_Scales_a_User's_Manual/links/5af580264585157136caee31/SF-36-Physical-and-Mental-Health-Summary-Scales-a-Users-Manual.pdf [accessed 2022-01-18]

44.   Wilson PM, Rogers WT, Rogers W, Wild TC. The psychological need satisfaction in exercise scale. J Sport Exerc Psychol 2006;28(3):231-251 [FREE Full text] [doi: 10.1123/jsep.28.3.231]

45.   Markland D, Tobin V. A modification to the behavioural regulation in exercise questionnaire to include an assessment of amotivation. J Sport Exerc Psychol 2004;26(2):191-196 [FREE Full text] [doi: 10.1123/jsep.26.2.191]

46.   Williams G, Freedman ZR, Deci EL. Supporting autonomy to motivate patients with diabetes for glucose control. Diabetes Care 1998 Oct;21(10):1644-1651. [doi: 10.2337/diacare.21.10.1644] [Medline: 9773724]

47.   McAuley E. The role of efficacy cognitions in the prediction of exercise behavior in middle-aged adults. J Behav Med 1992 Feb;15(1):65-88. [doi: 10.1007/bf00848378]

48.   Verplanken B, Orbell S. Reflections on past behavior: a self-report index of habit strength. J Appl Soc Psychol 2006;33(6):1313-1330. [doi: 10.1111/j.1559-1816.2003.tb01951.x]

49.   Ryan RM, Frederick C. On energy, personality, and health: subjective vitality as a dynamic reflection of well-being. J Pers 1997 Sep;65(3):529-565. [doi: 10.1111/j.1467-6494.1997.tb00326.x] [Medline: 9327588]

50.   Laird N. Further comparative analyses of pretest-posttest research designs. Am Statist 1983 Nov;37(4a):329-330. [doi: 10.1080/00031305.1983.10483133]

51.   Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials. Controll Clin Trials 2002 Dec;23(6):662-674. [doi: 10.1016/s0197-2456(02)00242-8]

52.   Efron B, Tibshirani RJ. An Introduction to the Bootstrap. New York: Chapman and Hall/CRC; 1994.

53.   Kelley K. The effects of nonnormal distributions on confidence intervals around the standardized mean difference: bootstrap and parametric confidence intervals. Edu Psychol Measur 2016 Jul 02;65(1):51-69. [doi: 10.1177/0013164404264850]

54.   Chen L, Peng CJ. The sensitivity of three methods to nonnormality and unequal variances in interval estimation of effect sizes. Behav Res Methods 2015 Mar 3;47(1):107-126. [doi: 10.3758/s13428-014-0461-3] [Medline: 24793819]

55.   Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York: Routledge; 1998.

56.   Clarke V, Braun V. Thematic analysis. In: Teo T, editor. Encyclopedia of Critical Psychology. New York: Springer; 2014.

57.   Rhodes RE, Dickau L. Moderators of the intention-behaviour relationship in the physical activity domain: a systematic review. Br J Sports Med 2013 Mar 25;47(4):215-225. [doi: 10.1136/bjsports-2011-090411] [Medline: 22278998]

58.   Sniehotta FF, Scholz U, Schwarzer R. Bridging the intention–behaviour gap: planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychol Health 2005 Apr;20(2):143-160. [doi: 10.1080/08870440512331317670]

59.   Harris T, Kerry SM, Limb ES, Furness C, Wahlich C, Victor CR, et al. Physical activity levels in adults and older adults 3-4 years after pedometer-based walking interventions: Long-term follow-up of participants from two randomised controlled trials in UK primary care. PLoS Med 2018 Mar 9;15(3):e1002526 [FREE Full text] [doi: 10.1371/journal.pmed.1002526] [Medline: 29522529]

60.   Wahlich C, Beighton C, Victor C, Normansell R, Cook D, Kerry S, et al. 'You started something … then I continued by myself': a qualitative study of physical activity maintenance. Prim Health Care Res Dev 2017 Aug 4;18(06):574-590. [doi: 10.1017/s1463423617000433]

61.   Normansell R, Smith J, Victor C, Cook DG, Kerry S, Iliffe S, et al. Numbers are not the whole story: a qualitative exploration of barriers and facilitators to increased physical activity in a primary care based walking intervention. BMC Public Health 2014 Dec 15;14(1):1272 [FREE Full text] [doi: 10.1186/1471-2458-14-1272] [Medline: 25511452]

62.   Lewis ZH, Pritting L, Picazo A, JeanMarie-Tucker M. The utility of wearable fitness trackers and implications for increased engagement: an exploratory, mixed methods observational study. Digit Health 2020 Jan 13;6:2055207619900059 [FREE Full text] [doi: 10.1177/2055207619900059] [Medline: 31976079]

63.   Kononova A, Li L, Kamp K, Bowen M, Rikard R, Cotten S, et al. The use of wearable activity trackers among older adults: focus group study of tracker perceptions, motivators, and barriers in the maintenance stage of behavior change. JMIR Mhealth Uhealth 2019 Apr 05;7(4):e9832 [FREE Full text] [doi: 10.2196/mhealth.9832] [Medline: 30950807]

64.   Peng W, Li L, Kononova A, Cotten S, Kamp K, Bowen M. Habit formation in wearable activity tracker use among older adults: qualitative study. JMIR Mhealth Uhealth 2021 Jan 19;9(1):e22488 [FREE Full text] [doi: 10.2196/22488] [Medline: 33464216]

65.   Bergman P. The number of repeated observations needed to estimate the habitual physical activity of an individual to a given level of precision. PLoS One 2018 Feb 1;13(2):e0192117 [FREE Full text] [doi: 10.1371/journal.pone.0192117] [Medline: 29390010]

66.   Gillison FB, Rouse P, Standage M, Sebire SJ, Ryan RM. A meta-analysis of techniques to promote motivation for health behaviour change from a self-determination theory perspective. Health Psychol Rev 2019 Mar 16;13(1):110-130. [doi: 10.1080/17437199.2018.1534071] [Medline: 30295176]

67.   Ntoumanis N, Ng JY, Prestwich A, Quested E, Hancox JE, Thøgersen-Ntoumani C, et al. A meta-analysis of self-determination theory-informed intervention studies in the health domain: effects on motivation, health behavior, physical, and psychological health. Health Psychol Rev 2021 Jun 03;15(2):214-244. [doi: 10.1080/17437199.2020.1718529] [Medline: 31983293]

68.   Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):112-118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]

Abbreviations

EQ: EuroQol
MET: metabolic equivalent of task

MIPACT: Multidimensional Individualized Physical Activity
MVPA: moderate to vigorous physical activity
PA: physical activity
PAL: physical activity level
RCT: randomized controlled trial
SDT: self-determination theory
SF-36: Short Form-36

XSL•FO
RenderX

Original Paper

eDOL mHealth App and Web Platform for Self-monitoring and Medical Follow-up of Patients With Chronic Pain: Observational Feasibility Study

Nicolas Kerckhove[1*], PhD; Noémie Delage[2*], MD; Sébastien Cambier[2], MSc; Nathalie Cantagrel[3], MD, PhD; Eric Serra[4], MD; Fabienne Marcaillou[2], MD; Caroline Maindet[5], MD; Pascale Picard[2], MD, PhD; Gaelle Martiné[6], MD; Rodrigue Deleens[7], MD; Anne-Priscille Trouvin[8], MD, PhD; Lauriane Fourel[9], MD; Gaelle Espagne-Dubreuilh[6], MD; Ludovic Douay[4], MD; Stéphane Foulon[4], MD; Bénédicte Dufraisse[6], MD; Christian Gov[10], MD; Eric Viel[11], MD, PhD; François Jedryka[11], MD; Sophie Pouplin[7], MD; Cécile Lestrade[3], MD; Emmanuel Combe[3], MD; Serge Perrot[8], MD, PhD; Dominique Perocheau[8], MD; Valentine De Brisson[9], MD; Pascale Vergne-Salle[6], MD, PhD; Patrick Mertens[10], MD, PhD; Bruno Pereira[2], PhD; Abdoul Jalil Djiberou Mahamadou[12], PhD; Violaine Antoine[12], PhD; Alice Corteval[13], PharmD; Alain Eschalier[13], MD, PharmD, PhD; Christian Dualé[2], MD; Nadine Attal[14], MD, PhD; Nicolas Authier[2], MD, PharmD, PhD

[1]Service de Pharmacologie médicale, Centre Hospitalier Universitaire de Clermont-Ferrand, Clermont-Ferrand, France

[2]Centre d'évaluation et de traitement de la douleur, Centre Hospitalier Universitaire de Clermont-Ferrand, Clermont-Ferrand, France

[3]Centre d'évaluation et de traitement de la douleur, Centre Hospitalier Universitaire de Toulouse, Toulouse, France

[4]Centre d'évaluation et de traitement de la douleur, Centre Hospitalier Universitaire d'Amiens, Amiens, France

[5]Centre d'évaluation et de traitement de la douleur, Centre Hospitalier Universitaire de Grenoble, Grenoble, France

[6]Centre d'évaluation et de traitement de la douleur, Centre Hospitalier Universitaire de Limoges, Limoges, France

[7]Centre d'évaluation et de traitement de la douleur, Centre Hospitalier Universitaire de Rouen, Rouen, France

[8]Centre d'évaluation et de traitement de la douleur, Assistance Publique–Hôpitaux de Paris Cochin, Paris, France

[9]Centre d'évaluation et de traitement de la douleur, Centre Hospitalier Régional de Bayeux, Bayeux, France

[10]Centre d'évaluation et de traitement de la douleur, Hospices Civils de Lyon Pierre Wertheimer, Lyon, France

[11]Centre d'évaluation et de traitement de la douleur, Centre Hospitalier Universitaire de Nîmes, Nîmes, France

[12]Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes, Université Clermont Auvergne, Aubière, France

[13]Institut Analgesia, Clermont-Ferrand, France

[14]Institut National de la Santé et de la Recherche Médicale, Unité 987 – Centre d'évaluation et de traitement de la douleur, Assistance Publique–Hôpitaux de Paris Ambroise Paré, Paris, France

[*]these authors contributed equally

Corresponding Author:
Nicolas Kerckhove, PhD
Service de Pharmacologie médicale
Centre Hospitalier Universitaire de Clermont-Ferrand
29 boulevard Joseph Girod
Clermont-Ferrand, 63000
France
Phone: 33 0473178414
Email: nkerckhove@chu-clermontferrand.fr

Abstract

Background: Chronic pain affects approximately 30% of the general population, severely degrades quality of life (especially in older adults) and professional life (inability or reduction in the ability to work and loss of employment), and leads to billions in additional health care costs. Moreover, available painkillers are old, with limited efficacy and can cause significant adverse effects. Thus, there is a need for innovation in the management of chronic pain. Better characterization of patients could help to identify the predictors of successful treatments, and thus, guide physicians in the initial choice of treatment and in the follow-up of their patients. Nevertheless, current assessments of patients with chronic pain provide only fragmentary data on painful daily

experiences. Real-life monitoring of subjective and objective markers of chronic pain using mobile health (mHealth) programs can address this issue.

Objective: We hypothesized that regular patient self-monitoring using an mHealth app would lead physicians to obtain deeper understanding and new insight into patients with chronic pain and that, for patients, regular self-monitoring using an mHealth app would play a positive therapeutic role and improve adherence to treatment. We aimed to evaluate the feasibility and acceptability of a new mHealth app called eDOL.

Methods: We conducted an observational study to assess the feasibility and acceptability of the eDOL tool. Patients completed several questionnaires using the tool over a period of 2 weeks and repeated assessments weekly over a period of 3 months. Physicians saw their patients at a follow-up visit that took place at least 3 months after the inclusion visit. A composite criterion of the acceptability and feasibility of the eDOL tool was calculated after the completion of study using satisfaction surveys from both patients and physicians.

Results: Data from 105 patients (of 133 who were included) were analyzed. The rate of adherence was 61.9% (65/105) after 3 months. The median acceptability score was 7 (out of 10) for both patients and physicians. There was a high rate of completion of the baseline questionnaires and assessments (mean 89.3%), and a low rate of completion of the follow-up questionnaires and assessments (63.8% (67/105) and 61.9% (65/105) respectively). We were also able to characterize subgroups of patients and determine a profile of those who adhered to eDOL. We obtained 4 clusters that differ from each other in their biopsychosocial characteristics. Cluster 4 corresponds to patients with more disabling chronic pain (daily impact and comorbidities) and vice versa for cluster 1.

Conclusions: This work demonstrates that eDOL is highly feasible and acceptable for both patients with chronic pain and their physicians. It also shows that such a tool can integrate many parameters to ensure the detailed characterization of patients for future research works and pain management.

Trial Registration: ClinicalTrial.gov NCT03931694; http://clinicaltrials.gov/ct2/show/NCT03931694

Introduction

Chronic pain affects approximately 30% of the general population [1-6] and was 1 of the top 5 leading causes of years lived with disability in 2016 [7], especially among older people [8]. Societal and economic issues are also crucial, as 60% of people with chronic pain are less able or unable to work, and 20% report having lost their job as a result [9]. The overall cost of chronic pain is estimated to be approximately €441 billion in Europe (equivalent to approximately US $496 billion) and $560 to $635 billion in the United States [10-12]. At the same time, the market for analgesic drugs represented approximately $68 billion in 2016, and an increase from 2% to 5% was forecast for 2021, with a further 5% increase by 2025 [13,14]. Unfortunately, available analgesics are old, their effectiveness is limited, with undesirable effects, and little progress has been made in recent years [15]. Thus, innovation is limited despite prolific basic research [16].

Various reasons are given for this, including the relevance of animal research [17]. In particular, because of the low success rate of validation of preclinical concepts during the transition to the clinic. Developments in this area could help progress, but such progress could also come from better patient characterization that would help to identify the predictors of successful treatments through research programs and enable physicians to carry out better decision-making regarding the initial choice of treatment and its follow-up. Subgroups of patients and criteria for response to particular treatments, for example, in patients with neuropathic pain [18], have been

identified; however, such characterization should not be limited to biomedical assessment but should also include biopsychosocial assessment. Moreover, current assessments of patients with chronic pain provide only fragmentary data on daily experiences because of recall bias. Thus, it is essential to modify the temporality in which patients' sensations are assessed, with real-life monitoring of subjective and objective markers of chronic pain. This strategy is currently being developed by several research teams evaluating smartphone apps or web platforms for use in managing the treatment of patients with chronic pain [19-24].

We hypothesized that regular self-monitoring by patients using a digital app would generate in-depth knowledge and new insights for physicians, and would allow patients to be active in their own care and benefit from web-based counseling. Regular self-monitoring would not only contribute to better patient characterization and help in choosing the most appropriate treatment but may also improve adherence to treatment. Moreover, recent studies [19,20,24-31] have highlighted the urgent need to develop eHealth self-monitoring programs for chronic pain and their therapeutic value—web-based pain management programs (The Pain Course) based on principles of cognitive behavior therapy were found to be beneficial for patients by reducing pain symptoms and associated comorbidities [20,32-34], and there is therapeutic interest in mobile health (mHealth) technologies for managing the medical treatments of patients suffering from chronic pain [27]. In this pilot study, we aimed to evaluate the feasibility and

XSL•FO

RenderX

acceptability of a new mHealth app and web platform, called eDOL, for patients and physicians.

Methods

Ethics

The study was approved by the *Comité de Protection des Personnes Ile de France V* (2018-A01790-5546) and is registered (NCT03931694). The study was conducted in accordance with French laws and regulations on research on human beings and data protection and with the Declaration of Helsinki [35].

Confidentiality and Data Entry and Processing

Data were collected and managed using the eDOL app, developed by Bepatient and hosted by Avenir Télématique. In accordance with the provisions relating to the confidentiality of information concerning, in particular, the people who took part in the research and the results obtained [36], individuals with direct access have taken all the necessary precautions to ensure the confidentiality of the information relating to the participants. These persons and the investigators themselves are subject to professional secrecy [37]. All data collected and transmitted to the sponsor (University Hospital of Clermont-Ferrand) were anonymized, and each patient had a single coded number. The head of research ensured that each patient was informed of which data were collected and that they did not object to their use or disclosure.

Answers to questionnaires and medical data were transmitted in spreadsheet format (Excel 2013, Microsoft Inc). All anonymized data were accessible to the biostatisticians (BP, SC, and AJD), the coordinator (ND), and the project manager (NK). Only the investigators could access their patients' personal data to identify them. A dashboard linking patients' identities and study IDs was available only on the investigators' professional interface on the eDOL web platform. The final database, used for statistical analyses, included only study IDs to preserve anonymity.

Study Design and Population

To evaluate the feasibility and acceptability of the eDOL app for the characterization, real-life monitoring of patients with chronic pain from 12 pain clinics in France took place between February 8, 2019 and January 8, 2020. The study was offered to all physicians in the investigating centers.

Participation in the study was offered to patients with chronic pain who did not have cancer, who were owners and regular users of a smartphone, and who were followed up in a pain clinic. All adult (≥18 years old) patients able to read and understand French and provide consent to participate in the study were included (with a yes-or-no choice on the eDOL app). Participants were free to withdraw their consent at any time by informing the sponsor. Each patient had access to the information document (paper or electronic) detailing the purpose, content, and conduct of the study. If they agreed to participate, they were asked to download the eDOL app and complete the questionnaires using the eDOL app. The URL to access this app was sent by email from physicians to their

patients. After downloading the app and creating their profile, patients could accept the general terms and conditions of use and confirm that they agree to the use of their medical data in this study.

Each patient had 1 initial study visit, during which, the physician introduced the study to the patient, checked their eligibility, explained the eDOL tool, and gave the patient a brief training document on how to use the eDOL smartphone app. Participants completed several questionnaires and assessments using the eDOL app over a period of 2 weeks (initial patient characterization) and then repeatedly over a period of 3 months and up to 6 months for patients who wished to continue using the app (weekly, quarterly, and half-yearly depending on the questionnaire). Physicians saw their patient at a follow-up visit that took place at least 3 months after the inclusion visit, with the possibility of continuing the follow-up for up to 6 months. The study was considered complete for patients who completed their questionnaires and assessments for at least 3 months and made a follow-up visit 3 to 6 months after the inclusion visit.

eDOL App

All data were collected using the eDOL digital health tool, which includes a smartphone app for patients that allows self-questionnaires and assessments to be completed for semiological monitoring (pain, anxiety, sleep quality), and a web interface for physicians, to allow them to graphically visualize the summary of data provided by their patients for clinical and therapeutic monitoring.

Patients completed questionnaires and weekly assessments (Multimedia Appendix 1). The questionnaires were divided into general questionnaires that were systematically filled in once only (sociodemographic, lifestyle and professional data, Pain Beliefs and Perceptions Inventory [38]; Evaluation of level of precariousness [39]; Injustice Experience Questionnaire [40]; Maslach Burn-out Inventory [41]; Toronto Alexithymia Scale [42]; Life Orientation Test-Revised [43]; Belief in a just world [44]; Job Content Questionnaire [45]; Big Five Inventory [46]) and questionnaires, assessing symptoms, comorbidities, and psychological and physiological states related to chronic pain, that were completed quarterly (Brief Pain Inventory [47]; Medical Outcomes Study Sleep Scale [48]) and, according to duration of follow-up, half-yearly (Tampa Scale of Kinesiophobia [49]; Pain Catastrophizing Scale [50]; Fear-avoidance beliefs [51]; EQ-5D-3L [52]; Hospital Anxiety Depression Scale [53]; Satisfaction With Life Scale [54]; Subjective Cognitive Complaints [55]). Some questionnaires were specific to a type of chronic pain (Neuropathic Pain Scale Inventory [56]; Western Ontario and McMaster Universities [57]; Rheumatoid Arthritis Impact of Disease [58]; Roland Morris Disability Questionnaire [59]; Irritable Bowel Severity Scoring System [60]; Fibromyalgia Impact Questionnaire [61]; Headache Impact Test [62]). Follow-up of patients (daily monitoring of various objective and subjective parameters related to the pathology), using assessments, was also integrated in the app, which allowed us to monitor the evolution of patients' pain and its repercussions. Assessments were in the form of an 11-point numeric rating scale (from 0 to 10), assessing the intensity of pain (average, minimum, or maximum intensity),

anxiety, fatigue, and the quality of sleep, morale, body comfort were assessed weekly for 3 to 6 months.

For physicians, the eDOL internet platform included a simple and ergonomic dashboard which allowed the physician to find all of their patients included in the study, with the following tabs: (1) Management, in which all of the medical records completed by the physician could be found (history, pain diagnosis, initial characterization, next appointment, consultation sheets and treatment sheets); (2) Health Measures, which showed a graphic display of the real-life follow-up of all the weekly assessments; and (3) Questionnaires, which showed all the questionnaires completed by the patients (display of questionnaire scores and answers to all the questions). The eDOL platform enabled physicians to complete medical elements during consultation visits with various medical form

(diagnosis, current treatments, examination results). The physicians could also activate new questionnaires to be filled in by their patients, either to complete the characterization (eg, specific questionnaires for pain diagnosis) or to evaluate other criteria (eg, evaluation of the Patients' Global Impression of Change after the introduction of a new treatment [63]). Diagnostic questionnaires (Posttraumatic stress disorder Checklist [64]; Neuropathic pain 4 [65]; Fibromyalgia Rapid Screening Tool [66]), reminders of the criteria for diagnoses (ROME IV for irritable bowel syndrome; Widespread pain index and Symptom severity scale of American college of rheumatology for fibromyalgia; Neuropathic Pain IASP Special Interest Groups for neuropathic pain), and screening tools for opioid misuse (Prescription Opioid Misuse Index [67] and Opioid Risk Tool [68]) were also at their disposal (Table 1).

Table 1. eDOL features.

Feature	Included in	Assessment point or interval	Details
Inclusion form	Investigator web platform	Initial visit	Last name, first name, email, ID number
		Initial visit	History (clinical, psychiatric, drug), clinical examination, medico-economic aspect (type of medical consultations), diagnosis of pain according to International Classification of Disease, 11th revision
Personal information	Smartphone app	Initial visit	Sociodemographic (work, alcohol use, tobacco use)
		Initial visit	Pain characterization: frequency, duration, aggravating and alleviating factors
Treatment forms	Investigator web platform	Updated at each consultation	Analgesics (name, dates, dosage, side effects); list of nonmedicinal techniques and other treatments (free text)
Assessments	Smartphone app	Repeated weekly	11-point numeric rating scale (0-10): sleep, morale, fatigue and energy, body comfort, anxiety, pain
Self-questionnaires	Smartphone app	During the first 2 weeks	5 sessions of questionnaires
		Not repeated	Fear-avoidance beliefs[a], Injustice Experience Questionnaire, Maslach Burn-out Inventory[a], Pain Beliefs and Perceptions Inventory, Evaluation of level of precariousness, Job Content Questionnaire[a], Life Orientation Test-Revised, Belief in a just world, Posttraumatic Stress Disorder Checklist[b], Toronto Alexithymia Scale Big Five Inventory
		Every 3 months	Fibromyalgia Impact Questionnaire[c], Headache Impact Test[c], irritable bowel severity scoring system[c], Prescription Opioid Misuse Index[b], Patients' Global Impression of Change[b], Neuropathic Pain Scale Inventory[b], Rheumatoid Arthritis Impact of Disease[b], Brief Pain Inventory, Medical Outcomes Study Sleep Scale
		Every 6 months	Tampa Scale of Kinesiophobia, Roland Morris Disability Questionnaire[c], Western Ontario and McMaster Universities[c], Pain Catastrophizing Scale, EuroQol 5 dimensions 3 levels, Hospital Anxiety Depression Scale, Satisfaction With Life Scale, Subjective Cognitive Complaints
Hetero-questionnaires	Investigator web platform	N/A[d]	Diagnostic validation: Neuropathic pain 4 + NEUPSIG (neuropathy), Widespread pain index and Symptom severity scale and Fibromyalgia Rapid Screening Tool (fibromyalgia), ROME IV (irritable bowel syndrome)
		Updated at each consultation	Others: Opioid Risk Tool
Consultation form	Investigator web platform	Updated at each consultation	clinical examination, medico-eco aspect, observance, benefit-risk ratio of treatments

[a]Work-related questionnaires.

[b]Optional questionnaires.

[c]Disease-specific questionnaires

[d]N/A: not applicable.

Study Outcomes

The primary study endpoint reflected the acceptability of the eDOL app and the feasibility of its use and was assessed with a satisfaction survey (based on the Patient Satisfaction Questionnaire Short Form [69] and the Client Satisfaction Questionnaire [70,71]) for patients (10 questions) and for participating physicians (12 questions) at the end of the study. The satisfaction survey (in French language) was sent to each patient 6 months after their inclusion visit and was sent to the physicians after the last patient follow-up, via the eDOL tool. Response options for each question ranged from 0 (strongly disagree with the statement) to 10 (strongly agree with the statement). A mean score of at least 7 out of 10 was considered to reflect satisfactory acceptability and feasibility of the eDOL

tool. The questionnaire completion rate and center participation (inclusion rate) were also calculated.

Secondary analyses to characterize participating patients, pain disorders, and related comorbidities, as well as clustering analysis of the participants to determine the profile determination of patients who adhered to the use of the app were undertaken to gain insight into the capabilities and added value of the tool for the characterization and the follow-up of patients with chronic pain.

Statistics

Sample Size

A minimum of 100 patients were to be included and analyzed. Such a large number of patients is quite satisfactory in terms of

descriptive analyses to study the feasibility of a multimodal eHealth tool. This number of patients is in line with that specified by Sundararaman et al [27] and those used in other recent studies [19,23,24,72] evaluating smartphone apps in patients with chronic pain. This number of patients (n=100) allowed us to analyze in an exploratory way: (1) the characterization of patients with chronic pain followed-up in pain clinics, (2) the description and understanding of their pain, (3) the multiple dimensions and the numerous neuropsychiatric repercussions of chronic pain, and (4) the clustering of the participants and the determination of adhering patients' profiles.

Statistical Analysis

We performed statistical analyses to determine if patients and physicians were satisfied with the tool and adhered to its use, and to identify interesting pain profiles of patients, and which profiles are most adherent (and for how long).

Patients were described according to epidemiological characteristics, clinical characteristics, and treatment characteristics. The key indicators for acceptability (patient and physician) were questionnaire completion and completion of follow-up medical forms. We determined the association between adherence and all baseline variables. A patient was defined as adherent if 100% of baseline questionnaires and 75% of assessments after 3 months follow-up were completed.

Continuous variables and scale variables (treated as ordinal data) were presented as mean and standard deviation (for normal distributions), or median and quartiles (for asymmetric distributions). The normality assumption was assessed with graphical criterion and the Shapiro-Wilk test. Categorical variables were expressed in number and percentage.

We performed clustering analysis. Patients were clustered according the symptoms and comorbidity information y. This clustering analysis included data imputation, principal component analysis of baseline data, and ascending hierarchical classification gathering 90% of total inertia. From these components, the hierarchical classification [73] in Euclidean coordinates was used.

Comparisons (baseline vs 3-month follow-up, by patient adherence, and by cluster) were performed using the chi-square test or Fisher exact test when assumptions to apply chi-square were not met (minimal level of expected number of cases under independence assumption), for categorical variables, and using analysis of variance (or Kruskal-Wallis tests when the assumptions to apply analysis of variance were not met). When the omnibus P value was statistically significant ($P<.05$), posthoc tests (independent t test or Mann-Whitney) were applied to compare subgroups with each other. The results were expressed using effect sizes (Cramer V for categorical data and eta square for quantitative data) with 95% confidence intervals. Pearson (preferred between all distributions acknowledged as Gaussian) or Spearman correlation coefficients (otherwise) were calculated depending on the nature of the distribution.

We used Stata (version 15, StataCorp LLC) and R (version 4.0.3) software. All statistical tests were 2-sided with type I error set at 5%.

Results

Study Population

Of 133 patients from 12 French pain clinics, 28 patients (28/133, 21.0%) did not install the eDOL app; data from 105 patients were analyzed. The first patient was enrolled on February 6, 2019, and the last patient was enrolled on October 31, 2019.

At baseline, participating patients were mostly middle-aged women, in a couple, nonsmoking, and professionals. Among these patients, 35.3% (30/85) were in work stoppage due to their chronic pain. A more detailed characterization of the patients, with the help of several validated questionnaires, mainly showed that a significant number were considered precarious (43.0%; 40/93), with kinesiophobia (72.0%; 67/93), alexithymia (51/100, 51%), degraded life satisfaction (51/92, 55.4%), catastrophism (47/100, 47.0%) and a possible cognitive disorder (77/93, 82.8%). More than 65% (63/94, 67.0%) of patients had impaired sleep, and 37.2% (35/94) and 27.7% (26/94) had proven anxiety or depressive disorders respectively.

Regarding the characterization of pain disorders and their treatments, most patients (76/83, 91.6%) had moderate to severe pain intensity, of which 20.5% (17/83) had a high chronic pain interference score (called "high impact chronic pain" [74]). Most patients (50/80, 62.5%) suffered from nociplastic pain, with a duration longer than 5 years for more than 50% (55/105, 52.4%) of patients. The majority of patients (56/105, 53.3%) described their chronic pain as permanent (with painful paroxysms every day and lasting >2 hours) and inducing frequent nocturnal awakenings (45/105, 42.8%). Finally, analgesic treatments used by the patients were mainly antidepressants followed by weak opioids (with or without paracetamol), and antiepileptics to a lesser extent. In parallel, 89.2% (66/74) of patients used nonmedicinal analgesic treatments.

There was no difference in any of these characteristics between baseline and the 3-month follow-up (Multimedia Appendix 2).

Primary Objective: Feasibility and Acceptability

Among 105 patients, 65 (61.9%) adhered to the use of the eDOL tool and 50 patients continued using the eDOL tool up to 6-month follow-up (Figure 1).

In detail, the overall rate of patient who completed the baseline questionnaires was 89.3% (range 79.0%-95.2%). The quarterly questionnaires, Brief Pain Inventory and Medical Outcomes Study Sleep Scale, were repeatedly filled at 3-month follow-up by 63.8% (67/105) of patients. For the half-yearly questionnaires (Tampa Scale of Kinesiophobia; Pain Catastrophizing Scale; EQ-5D-3L; Hospital Anxiety Depression Scale; Satisfaction With Life Scale and Subjective Cognitive Complaints), 58.7% (range 53.8%-63.1%) of patients completed the questionnaires. The filling rate of the weekly assessments for the real-life monitoring of the different parameters (pain, moral, anxiety, fatigue, sleep and body comfort) was 88.6% (93/105) of patients at the end of the first week and 61.9% (65/105) at 3-month follow-up (Table 2; Figure 2; Multimedia Appendix 3). Due to the small number of patients, we did not show the results concerning the specific questionnaires, filled by only a few

patients according to their professional situation (questionnaires on work) and their type of pain (disease-specific questionnaires). The rate of patients whose various medical follow-up forms were completed by the investigators (inclusion, treatment and consultation) was 70.7% (range 62.9-76.2%) (Table 2).

Figure 1. Study flowchart.
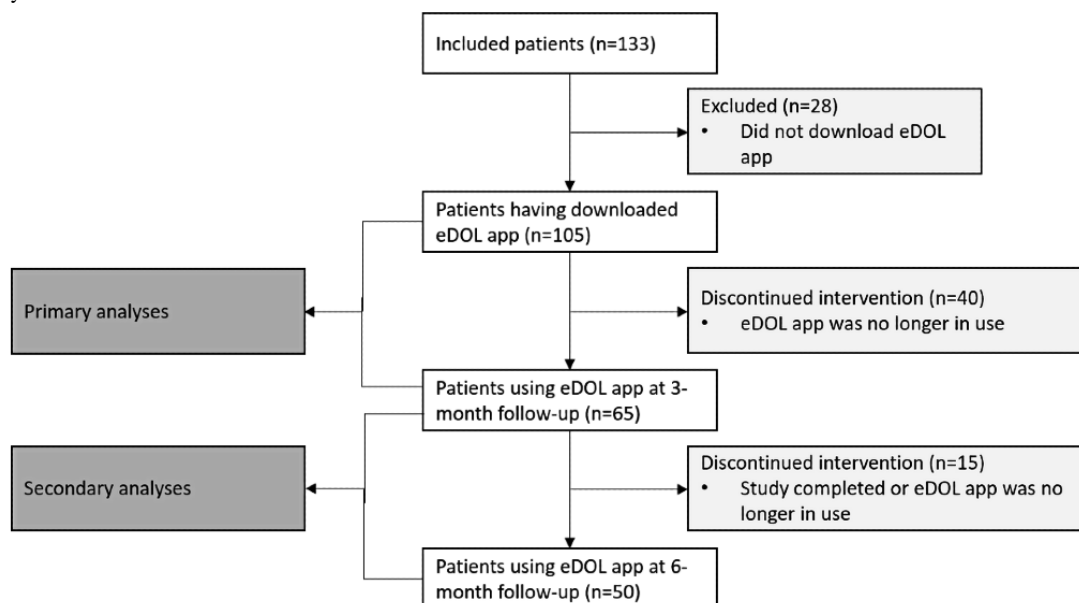
Table 2. Questionnaire completion.

Assessment	Baseline (n=105), n (%)	3-month follow-up (n=105), n (%)	6-month follow-up (n=65), n (%)
Physician baseline and follow-up forms			
Inclusion form (baseline)	77 (73.3)	N/A[a]	N/A
Diagnosis form (baseline)	80 (76.2)	N/A	N/A
Treatment form (baseline and follow-up)	74 (70.5)	N/A	N/A
Consultation form (follow-up)	66 (62.9)	N/A	N/A
Self-administered questionnaires and assessments			
Weekly assessments	93 (88.6)	65 (61.9)	50 (76.9)
Toronto Alexithymia Scale	100 (95.2)	N/A	N/A
Injustice Experience Questionnaire	100 (95.2)	N/A	N/A
Pain Beliefs and Perceptions Inventory	92 (87.6)	N/A	N/A
Life Orientation Test-Revised	94 (89.5)	N/A	N/A
Belief in a just world	94 (89.5)	N/A	N/A
Evaluation of level of precariousness	93 (88.6)	N/A	N/A
Big Five Inventory	92 (87.6)	N/A	N/A
MOS-Sleep Scale	94 (89.5)	67 (63.8)	39 (60.0)
Brief Pain Inventory	93 (88.6)	67 (63.8)	38 (58.5)
Pain Catastrophizing Scale	100 (95.2)	N/A	40 (61.5)
Satisfaction With Life Scale	92 (87.6)	N/A	35 (53.8)
Subjective Cognitive Complaints	93 (88.6)	N/A	35 (53.8)
EQ-5D-3L	83 (79.0)	N/A	36 (55.4)
Hospital Anxiety Depression Scale	94 (89.5)	N/A	41 (63.1)
Tampa Scale of Kinesiophobia	93 (88.6)	N/A	41 (63.1)

[a]N/A: not applicable.

Figure 2. Completion rate over time.
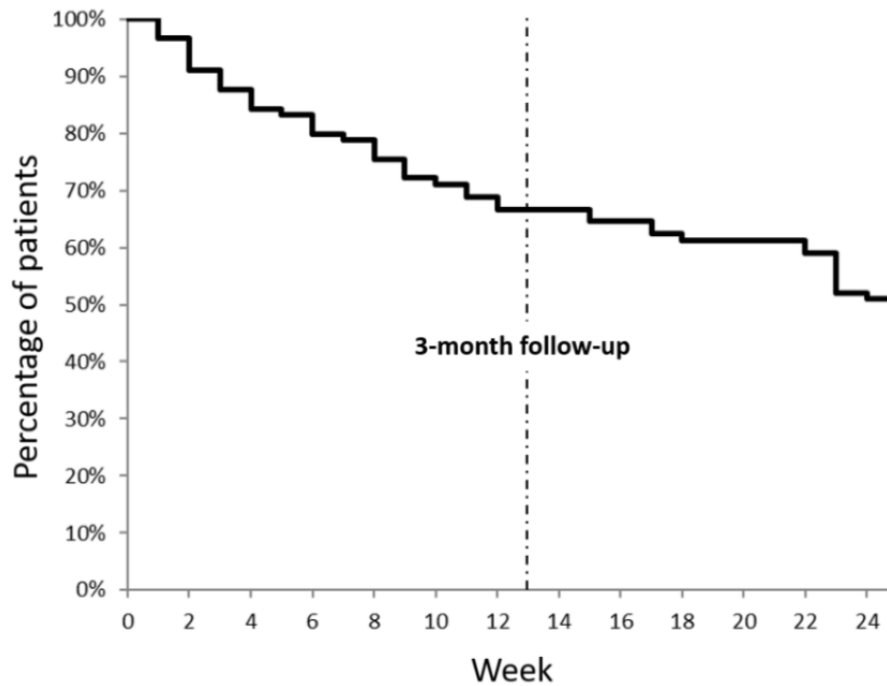
Among the 12 pain clinics participating in the study, 10 (83.3%) included patients, and 2 withdrew from participation before the start of the study. The median inclusion number per center was 8 (IQR 5.0, 14.0) patients. The inclusion objective (at least 100 analyzable patients) was achieved in less than a year as requested from the investigating centers.

The satisfaction questionnaire was filled in by 65.7% (69/105) of patients at the end of the study. The median acceptability score was 7.0 (IQR 6.1, 7.6), with only 9.5% (10/105) of the patients providing a rating less than 5.0 out of 10. Moreover, 88.6% (93/105) of the patients who responded wanted to participate in the further development of the eDOL app. The items with the lowest scores corresponded to the patients' perception of the physicians' use of eDOL in their follow-up (mean 5.7, SD 3.1), patients' perception of the potential positive

impact of eDOL on their pain management (mean 5.8, SD 2.7), and quality of life (mean 5.6, SD 2.4).

A total of 21 physicians participated in the study and included at least one patient, and 15 (71.4%) answered the satisfaction questionnaire. The physicians were mostly women (14/21, 66.7%), approximately 50.1 years old (range 33-61), and were from various specialties (2 neurologists, 2 psychiatrists, 3 anesthesiologists, 3 rheumatologists, and 5 general practitioners). The median acceptability score was 7.2 (IQR 6.8, 8.3), with only 6.7% (1/15) of physicians rating less than 5.0 out of 10. The items with the lowest scores corresponded to the compatibility of eDOL with the electronic medical file systems (mean 5.0, SD 2.3) and the possibility of eventually replacing the electronic medical files with the eDOL tool (mean 4.4, SD 1.9) (Table 3).

Table 3. Physician and patient acceptability of eDOL.

Acceptability questionnaire	Score (out of 10), mean (SD)
Physician	
The training and support provided was sufficient to use eDOL correctly	7.3 (1.4)
After the first training session, it is easy to use eDOL on a daily basis	6.9 (2.3)
The technical support (email and phone) was available to assist me if needed	8.3 (1.2)
eDOL offers questionnaires and assessments adapted to the multidimensional characterization of my patients	8.3 (1.2)
The forms I had to fill in for each patient are adapted and they correspond to the information I usually collect	6.8 (2.0)
Thanks to the export function provided in eDOL, I was able to retrieve the completed information for my patients. I was then able to print it (for my patient records) and/or import it into my hospital's electronic management system	5.0 (2.3)
The eDOL platform is complete enough to be able to replace my medical records one day	4.4 (1.9)
I would like to continue using eDOL in the future	7.3 (2.0)
eDOL will be useful in my daily medical practice	6.8 (1.6)
eDOL will allow me to better monitor my patients to improve their care	7.1 (1.6)
eDOL will be useful for developing clinical research on pain (creation of an e-cohort of patients with chronic pain)	9.0 (0.9)
eDOL will be useful for the clinical research projects conducted by my pain clinic	8.5 (1.7)
Patient[a]	
After reading the explanatory document provided by the physician, it was easy for me to use eDOL	8.4 (2.1)
After the first use, it is easy to use eDOL on a daily basis	8.7 (1.9)
The technical support was responsive enough when I asked for it	7.0 (2.7)
eDOL offers questionnaires and assessments that I feel are suitable for monitoring my pain and its impact on my daily life	7.0 (2.1)
I believe that the information I have entered in eDOL allows my doctor to better understand my pain and improve its management	6.9 (2.5)
During the time that I have been using eDOL, I feel that my doctor has better monitored my symptoms and that my pain has been better managed	5.7 (3.1)
I believe that the information I have entered in eDOL will also help researchers to better understand chronic pain and to identify new avenues of research	7.5 (2.3)
I think that eDOL will help me in my daily life to better manage my pain and its impact on my daily life	5.8 (2.7)
I think that eDOL will gradually improve my quality of life	5.6 (2.4)
I would like to continue using eDOL in the future	7.6 (2.8)

[a]88.5% indicated they would participate in the next phase of study on the new version of eDOL.

Secondary Objectives

We obtained 4 clusters that did not differ with respect to sociodemographic and chronic pain characteristics (except for pain interference with daily life) and their treatments (Multimedia Appendix 4). Interestingly, all patient characteristics obtained from validated biopsychosocial questionnaires differed between profiles. In particular, the patients in cluster 4 had more severe scores in various biopsychosocial and comorbidity scales (precariousness, anxiety, depression, kinesiophobia, sleep and cognitive disorders; $P<.001$) associated with a greater impact of pain and conversely for cluster 1. Clusters 2 and 3 were intermediate groups.

In Cluster 4, 80.0% (24/30) of patients adhered to the use of the tool, compared with 51.0% (19/37), 64.3% (9/14), and 43.5% (10/23) in clusters 1, 2, and 3, respectively (Multimedia Appendix 4). Moreover, type of pain was also related to adherence, with patients suffering from nociplastic pain who seemed to be more adherent than others (30/45, 66.7%; $P=.01$). It is noteworthy that 2 other items (presence of cognitive disorders and alexithymia) were related to adherence ($P=.04$), but with a small effect size (Cramer $V=.03$ and Cramer $V=.20$ respectively).

With reference to the profile of patients in cluster 4, the most severe patients, with a significant impact of pain on their daily life ($P=.03$), seemed to be those who adhered most to eDOL (24/62, 38.7% of patients who adhered to the use of the tool were in cluster 4).

Discussion

Overview

As this was primarily a feasibility study, we first discuss considerations regarding the data collection and acceptability, and then our exploratory results with respect to conducting future works and improving eDOL. Because of the low number of patients (and thus the limited longitudinal outcome data collected), we did not explore the impact of eDOL on pain disorders and related comorbidities.

Feasibility

Our results showed a rate of adherence, after 3-month follow-up, of approximately 60% (65/105, 61.9%) of patients using eDOL. Three similar recent studies [24,27,31], which assessed a smartphone app that enables patients with chronic pain to assess, monitor, and communicate their status to their providers, showed that 76%, 70%, and 72% of patients used the app for 3 months. Another study [20], which assessed a remotely delivered pain management program in a web-based format (web platform), showed that 76% of patients adhered [20] for at least 3 months. A study [75] with adolescents with chronic pain showed a high level of adherence (78%) and satisfaction, and a study [72] with patients with multiple sclerosis and migraine that evaluated the feasibility of using a smartphone app for patient follow-up showed an adherence rate of 49% after 90 days. The eDOL tool seems to be accepted in a similar way to these other smartphone-based or web-based apps. In our study, only an email reminder was sent to our patients if they had not used the app within 2 weeks after their inclusion and only 1 visit (included in their usual care path) was scheduled after at least 3 months. The studies [20,24,27,31,72,75] cited above included regular telephone follow-up or frequent visits. Moreover, according to the mean score (5.7, SD 3.1) for the statement "During the time that I have been using eDOL, I feel that my doctor has better monitored my symptoms and that my pain has been better managed," patients perceived that there was a lack of involvement of physicians in the eDOL tool. A study [76] showed that strong involvement by physicians increases adherence and the effectiveness of eHealth tools. Therefore, we can assume that a closer relationship with our patients (medical follow-up rhythm and involvement of physicians) would have further increased their adherence. This is undoubtedly a direction of research that should be taken for the future use of the app and patient follow-up; however, we must keep in mind that the aim of a real-life eHealth app is to be of little or no constraint for patients and to improve their medical follow-up, while lightening the physician's workload.

The good acceptability score, from both patients and physicians, reflects the interest expressed for eDOL and its contribution to the follow-up. Thus, eDOL could meet the urgent need to develop self-management and chronic pain management strategies through eHealth programs (internet, smartphone apps), and their therapeutic interest, as described by several studies [19,20,24-30].

Exploratory Analyses

In our exploratory analyses, our study population was similar to the profile of patients suffering from chronic pain in France [77], Germany [78], the United Kingdom [2], Canada [79], or the United States [74,80,81]—predominantly female, middle-aged, active population of lower socioeconomic status (precariousness, employment status, level of education), with pain lasting more than 5 years and suffering from psychological distress and from fairly severe chronic pain that has a significant impact on their lives (92% with moderate-to-severe pain, 20% with high impact chronic pain [74] and 43% with sleep disorders, such as awakenings due to pain at least once a night), mainly treated by antidepressants, and weak opioids. Interestingly, most did not simultaneously explore sociodemographic, psychological, pain disorders, and treatments characteristics.

With our smartphone app, we were able to collect data on precariousness, kinesiophobia, catastrophism, alexithymia, feelings of injustice, personality, life satisfaction, beliefs about pain, anxiety-depression, sleep, quality of life, cognitive disorders, optimism and belief in a just world. We made this choice because all of these factors are related to chronic pain [38,40,50,77,82-90] and we wanted to evaluate the ability of eDOL to characterize our patients precisely. Thus, the strength of eDOL is that it enables the integration of a large panel of validated questionnaires that, in turn, enable the precise characterization of the patients, especially regarding their emotional and psychological state, chronic pain, and related comorbidities. This characterization will eventually provide a large amount of data for care and research, and rely on a multimodal exploratory analysis of the determinants and repercussions of chronic pain, and their evolution in a real-life context, taking into account all the environmental events likely to influence chronic pain (treatments, history, comorbidities).

Finally, the multifactorial analysis of all our data enabled us to group our study population into 4 clusters. Interestingly, subpopulations of our patients could be distinguished only on the basis of biopsychosocial questionnaires and impact of pain on daily life whereas sociodemographic aspects, symptomatology, seniority and treatment of pain did not differ between our clusters. Cluster 4 represented patients with more disabling chronic pain, more severe comorbidities, and more pronounced psychological disorders, while cluster 1 represented patients with chronic pain that has little impact on their daily life, as well as a lower presence of comorbidity. Cluster 4 had a higher proportion of adherent patients. Our findings were similar results to those in a recent study [31], which showed that adherent patients correspond to patients with high impact chronic pain. These results seem consistent because patients with high impact chronic pain [74] and associated comorbidities are more in need of a tool that potentially improves their medical follow-up and are therefore more inclined to use it. Moreover, nociplastic pain was related to adherence (P=.01). According to our experience with chronic pain treatment management, this characteristic could be explained both by the fact that patients suffering from nociplastic pain (especially fibromyalgia) are younger than the general chronic pain population (and thus, more digital friendly) and very involved in the management of

their pain. Interestingly, the presence of cognitive disorders and alexithymia, independent of clusters, was related to adherence (*P*=.04). We hypothesize that patients with these disorders are aware of this and compensate by using eDOL as a digital companion, resulting in better adherence.

In addition, our results support the importance of questionnaires assessing the biopsychosocial aspect of chronic pain in addition to the biomedical aspect in the medical follow-up and characterization of patients with chronic pain. Moreover, in a classical medical follow-up, patients typically only see their pain specialist every 3 to 6 months. During these interviews, patients often have difficulties recalling their various symptoms and the impact of their pain over the past few months, which corresponds to a recall or memory bias [91]. Nevertheless, a review [92] demonstrates that the results of previous studies investigating this topic are highly variable. Some studies have shown that pain is remembered accurately [93-95], but others highlighted that patients tend to overestimate [96,97] or underestimate their pain [98]. Thus, a definitive answer to this question is still lacking, but real-life monitoring of different biopsychosocial and biomedical factors related to pain (not only pain intensity), using digital tools such as eDOL, could be a benefit in treatment management and the follow-up of patients.

Limitations

There was a selection bias mainly because requiring the use of a smartphones excludes patients who do not have or do not know how to use this tool. This could exclude the older or more precarious patients. Nevertheless, in view of our results, the age of the participants and the rate of precariousness were similar to those found in the general French population, with and without chronic pain [77,99]. We also observed that our population included many patients with nociplastic pain (mainly fibromyalgia, 50/80, 62.5%), which was not the case in other foreign studies [20,24,27,31,72,75]. Another French study [77] also found a high rate of fibromyalgia (42%), which seems to show that the population of French pain clinics includes a large proportion of fibromyalgia patients. Thus, we can conclude that this bias has little impact on our results. The second limitation was a measurement bias, which occurs frequently in observational studies [100]. Nevertheless, self-reporting permits a wider range of responses than many other data collection designs [101]. Measurement bias can arise from recall period, selective recall, social desirability, or sampling approach. In our study, the recall period might be the major risk [100]. Since all the questions dealt with the present moment or, at the latest, 1 to 2 weeks earlier, the recall bias can be considered negligible.

Moreover, our satisfaction survey was not a standardized but was a custom-made tool. We built this tool based on existing tools, such as the Patient Satisfaction Questionnaire [69] and the Client Satisfaction Questionnaire [70,71], and adapted it to our study and to the eDOL tool so that we could have specific feedback for improvement. It should be noted that the tools on which ours were based have little or no relevance to mHealth interventions [70], hence the need to create one adapted specifically for our study.

Finally, only physicians were involved in this feasibility study; other members of the care team, such as nurses, physiotherapists, and psychologists, did not participate in the study. The absence of the point of view of the rest of care teams is a limitation to the interpretation of the acceptability of the eDOL tool. In future studies of the eDOL tool, we plan to include all the members of the care team as well as the addition of a chatbot and a new therapeutic education tool.

Conclusions

The study demonstrated the feasibility and acceptability of eDOL for both patients with chronic pain and their physicians. These points justify continuing the deployment of the tool while providing information to improve its use and adherence to provide patients with chronic pain and their physicians with a better longitudinal characterization of pain and its impacts for an optimized and more personalized therapeutic management.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient and physician satisfaction questionnaires.
[DOC File , 44 KB - formative_v6i3e30052_app1.doc]

Multimedia Appendix 2
eDOL features.
[DOC File , 53 KB - formative_v6i3e30052_app2.doc]

Multimedia Appendix 3
Questionnaire completion data.
[DOC File , 42 KB - formative_v6i3e30052_app3.doc]

XSL•FO
RenderX

Multimedia Appendix 4
Study population details.
[DOC File , 377 KB - formative_v6i3e30052_app4.doc]

References

1. Elliott AM, Smith BH, Penny KI, Smith WC, Chambers WA. The epidemiology of chronic pain in the community. Lancet 1999 Oct 09;354(9186):1248-1252. [doi: 10.1016/s0140-6736(99)03057-3] [Medline: 10520633]

2. van Hecke O, Torrance N, Smith BH. Chronic pain epidemiology and its clinical relevance. Br J Anaesth 2013 Jul;111(1):13-18 [FREE Full text] [doi: 10.1093/bja/aet123] [Medline: 23794640]

3. Docking RE, Fleming J, Brayne C, Zhao J, Macfarlane GJ, Jones GT, Cambridge City over-75s Cohort Study collaboration. Epidemiology of back pain in older adults: prevalence and risk factors for back pain onset. Rheumatology (Oxford) 2011 Sep;50(9):1645-1653. [doi: 10.1093/rheumatology/ker175] [Medline: 21606130]

4. Thomas E, Peat G, Harris L, Wilkie R, Croft PR. The prevalence of pain and pain interference in a general population of older adults: cross-sectional findings from the North Staffordshire Osteoarthritis Project (NorStOP). Pain 2004 Jul;110(1-2):361-368. [doi: 10.1016/j.pain.2004.04.017] [Medline: 15275787]

5. Gobina I, Villberg J, Välimaa R, Tynjälä J, Whitehead R, Cosma A, et al. Prevalence of self-reported chronic pain among adolescents: evidence from 42 countries and regions. Eur J Pain 2019 Feb;23(2):316-326. [doi: 10.1002/ejp.1306] [Medline: 30098106]

6. Larsson C, Hansson EE, Sundquist K, Jakobsson U. Chronic pain in older adults: prevalence, incidence, and risk factors. Scand J Rheumatol 2017 Jul;46(4):317-325. [doi: 10.1080/03009742.2016.1218543] [Medline: 27885914]

7. GBD 2016 DiseaseInjury IncidencePrevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017 Sep 16;390(10100):1211-1259 [FREE Full text] [doi: 10.1016/S0140-6736(17)32154-2] [Medline: 28919117]

8. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain 2006 May 10;10(4):287-287. [doi: 10.1016/j.ejpain.2005.06.009]

9. Attal N, Lanteri-Minet M, Laurent B, Fermanian J, Bouhassira D. The specific disease burden of neuropathic pain: results of a French nationwide survey. Pain 2011 Dec;152(12):2836-2843. [doi: 10.1016/j.pain.2011.09.014] [Medline: 22019149]

10. Impact of pain on society costs the EU up to 441 billion Euros annually. Societal Impact of Pain (SIP). URL: https://www.sip-platform.eu/press-area/article/impact-of-pain-on-society-costs-the-eu-up-to-441-billion-euros-annually [accessed 2019-10-30]

11. Gaskin DJ, Richard P. The economic costs of pain in the United States. J Pain 2012 Aug;13(8):715-724. [doi: 10.1016/j.jpain.2012.03.009] [Medline: 22607834]

12. Breivik H, Eisenberg E, O'Brien T, OPENMinds. The individual and societal burden of chronic pain in Europe: the case for strategic prioritisation and action to improve knowledge and availability of appropriate care. BMC Public Health 2013;13:1229 [FREE Full text] [doi: 10.1186/1471-2458-13-1229] [Medline: 24365383]

13. Global medicine spending and usage trends: outlook to 2025. The IQVIA Institute. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/global-medicine-spending-and-usage-trends-outlook-to-2025 [accessed 2021-04-29]

14. The IQVIA Institute. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/reports-archive [accessed 2021-04-29]

15. Crofford LJ. Adverse effects of chronic opioid therapy for chronic musculoskeletal pain. Nat Rev Rheumatol 2010 Apr;6(4):191-197. [doi: 10.1038/nrrheum.2010.24] [Medline: 20357788]

16. Mogil JS. Animal models of pain: progress and challenges. Nat Rev Neurosci 2009 Apr;10(4):283-294. [doi: 10.1038/nrn2606] [Medline: 19259101]

17. Mouraux A, Bannister K, Becker S, Finn DP, Pickering G, Pogatzki-Zahn E, et al. Challenges and opportunities in translational pain research - an opinion paper of the Working Group on Translational Pain Research of the European Pain Federation (EFIC). Eur J Pain 2021 Apr;25(4):731-756. [doi: 10.1002/ejp.1730] [Medline: 33625769]

18. Bouhassira D, Branders S, Attal N, Fernandes AM, Demolle D, Barbour J, et al. Stratification of patients based on the neuropathic pain symptom inventory: development and validation of a new algorithm. Pain 2021 Apr 01;162(4):1038-1046. [doi: 10.1097/j.pain.0000000000002130] [Medline: 33136982]

19. Jamison RN, Jurcik DC, Edwards RR, Huang C, Ross EL. A pilot comparison of a smartphone app with or without 2-way messaging among chronic pain patients: who benefits from a pain app? Clin J Pain 2017 Aug;33(8):676-686 [FREE Full text] [doi: 10.1097/AJP.0000000000000455] [Medline: 27898460]

20. Manjula BN, Acharya AS, Vithayathil PJ. Deamidated active intermediates in the irreversible acid denaturation of ribonuclease-A. Int J Pept Protein Res 1976;8(3):275-282. [doi: 10.1111/j.1399-3011.1976.tb02504.x] [Medline: 6396]

21. Suso-Ribera C, Castilla D, Zaragozá I, Ribera-Canudas MV, Botella C, García-Palacios A. Validity, reliability, feasibility, and usefulness of pain monitor: a multidimensional smartphone app for daily monitoring of adults with heterogenous chronic pain. Clin J Pain 2018 Oct;34(10):900-908. [doi: 10.1097/AJP.0000000000000618] [Medline: 29659375]

22. Shadd JD, Ryan BL, Maddocks HL, McKay SD, Moulin DE. Neuropathic pain in a primary care electronic health record database. Eur J Pain 2015 May;19(5):715-721. [doi: 10.1002/ejp.594] [Medline: 25504680]

23. Minen MT, Jalloh A, Ortega E, Powers SW, Sevick MA, Lipton RB. User design and experience preferences in a novel smartphone application for migraine management: a think aloud study of the RELAXaHEAD application. Pain Med 2019 Feb 01;20(2):369-377 [FREE Full text] [doi: 10.1093/pm/pny080] [Medline: 29868895]

24. Jamison RN, Mei A, Ross EL. Longitudinal trial of a smartphone pain application for chronic pain patients: predictors of compliance and satisfaction. J Telemed Telecare 2018 Feb;24(2):93-100. [doi: 10.1177/1357633X16679049] [Medline: 27831496]

25. Gogovor A, Visca R, Auger C, Bouvrette-Leblanc L, Symeonidis I, Poissant L, et al. Informing the development of an internet-based chronic pain self-management program. Int J Med Inform 2017 Jan;97:109-119. [doi: 10.1016/j.ijmedinf.2016.10.005] [Medline: 27919369]

26. McGuire BE, Henderson EM, McGrath PJ. Translating e-pain research into patient care. Pain 2017 Feb 10;158(2):190-193. [doi: 10.1097/j.pain.0000000000000686] [Medline: 27525833]

27. Sundararaman LV, Edwards RR, Ross EL, Jamison RN. Integration of mobile health technology in the treatment of chronic pain. Reg Anesth Pain Med 2017;42(4):488-498. [doi: 10.1097/aap.0000000000000621]

28. Martin CL, Bakker CJ, Breth MS, Gao G, Lee K, Lee MA, et al. The efficacy of mobile health interventions used to manage acute or chronic pain: a systematic review. Res Nurs Health 2021 Feb;44(1):111-128. [doi: 10.1002/nur.22097] [Medline: 33341989]

29. Suso-Ribera C, Castilla D, Zaragozá I, Mesas, Server A, Medel J, et al. Telemonitoring in chronic pain management using smartphone apps: a randomized controlled trial comparing usual assessment against app-based monitoring with and without clinical alarms. Int J Environ Res Public Health 2020 Sep 09;17(18) [FREE Full text] [doi: 10.3390/ijerph17186568] [Medline: 32916983]

30. Mariano TY, Wan L, Edwards RR, Lazaridou A, Ross EL, Jamison RN. Online group pain management for chronic pain: preliminary results of a novel treatment approach to teletherapy. J Telemed Telecare 2021 May;27(4):209-216. [doi: 10.1177/1357633X19870369] [Medline: 31431133]

31. Ross EL, Jamison RN, Nicholls L, Perry BM, Nolen KD. Clinical integration of a smartphone app for patients with chronic pain: retrospective analysis of predictors of benefits and patient engagement between clinic visits. J Med Internet Res 2020 Apr 16;22(4):e16939 [FREE Full text] [doi: 10.2196/16939] [Medline: 32297871]

32. Dear BF, Courtney C, Khor KE, McDonald S, Ricciardi T, Gandy M, et al. The pain course: exploring the feasibility of an internet-delivered pain management program when offered by a tertiary pain management service. Clin J Pain 2018 Jul;34(6):505-514. [doi: 10.1097/AJP.0000000000000565] [Medline: 29077622]

33. Dear BF, Gandy M, Karin E, Fogliati R, Fogliati VJ, Staples LG, et al. The pain course: 12- and 24-month outcomes from a randomized controlled trial of an internet-delivered pain management program provided with different levels of clinician support. J Pain 2018 Dec;19(12):1491-1503. [doi: 10.1016/j.jpain.2018.07.005] [Medline: 30099209]

34. Dear BF, Gandy M, Karin E, Staples LG, Johnston L, Fogliati VJ, et al. The pain course: a randomised controlled trial examining an internet-delivered pain management program when provided with different levels of clinician support. Pain 2015 Oct;156(10):1920-1935 [FREE Full text] [doi: 10.1097/j.pain.0000000000000251] [Medline: 26039902]

35. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]

36. Code de la santé publique article R5121-13. Légifrance. URL: https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000025788178 [accessed 2022-02-06]

37. Code pénal article 226-13. Légifrance. URL: https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000006417945/ [accessed 2022-02-06]

38. Williams DA, Thorn BE. An empirical assessment of pain beliefs. Pain 1989 Mar;36(3):351-358. [doi: 10.1016/0304-3959(89)90095-X] [Medline: 2710564]

39. Sass C, Moulin JJ, Gueguen R, Abric L, Dauphinot V, Dupre C, et al. Le score Epices?: un score individuel de précarité construction du score et mesure des relations avec des données de santé, dans une population de 197 389 personnes. Bulletin Epidémiologique Hebdomadaire 2006 Apr 4;14:93-96 [FREE Full text]

40. Sullivan MJL, Adams H, Horan S, Maher D, Boland D, Gross R. The role of perceived injustice in the experience of chronic pain and disability: scale development and validation. J Occup Rehabil 2008 Sep;18(3):249-261. [doi: 10.1007/s10926-008-9140-5] [Medline: 18536983]

41. Maslach C, Jackson SE. The measurement of experienced burnout. J Organiz Behav 1981 Apr;2(2):99-113. [doi: 10.1002/job.4030020205]

42. Bagby RM, Taylor GJ, Parker JD. The twenty-item Toronto Alexithymia Scale--II. convergent, discriminant, and concurrent validity. J Psychosom Res 1994 Jan;38(1):33-40. [doi: 10.1016/0022-3999(94)90006-x] [Medline: 8126688]

43. Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. J Pers Soc Psychol 1994 Dec;67(6):1063-1078. [Medline: 7815302]

44. Lucas T, Zhdanova L, Alexander S. Procedural and distributive justice beliefs for self and others. J Individ Differ 2011 Jan;32(1):14-25. [doi: 10.1027/1614-0001/a000032]

45. Karasek R, Brisson C, Kawakami N, Houtman I, Bongers P, Amick B. The Job Content Questionnaire (JCQ): an instrument for internationally comparative assessments of psychosocial job characteristics. J Occup Health Psychol 1998 Oct;3(4):322-355. [Medline: 9805280]

46. John O, Srivastava S. The Big Five Trait taxonomy: history, measurement, and theoretical perspectives. In: Handbook of Personality: Theory and Research 2nd ed. New York, NY: Guilford Press; 1999:102-138.

47. Cleeland CS, Ryan KM. Pain assessment: global use of the brief pain inventory. Ann Acad Med Singapore 1994 Mar;23(2):129-138. [Medline: 8080219]

48. Hays RD, Martin SA, Sesti AM, Spritzer KL. Psychometric properties of the Medical Outcomes Study sleep measure. Sleep Med 2005 Jan;6(1):41-44. [doi: 10.1016/j.sleep.2004.07.006] [Medline: 15680294]

49. Miller RP, Kori SH, Todd DD. The Tampa scale. Clin J Pain 1991;7(1):51. [doi: 10.1097/00002508-199103000-00053]

50. Sullivan M, Stanish W, Waite H, Sullivan M, Tripp D. Catastrophizing, pain, and disability in patients with soft-tissue injuries. Pain 1998 Sep;77(3):253-260. [doi: 10.1016/S0304-3959(98)00097-9] [Medline: 9808350]

51. Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A fear-avoidance beliefs questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain 1993 Feb;52(2):157-168. [Medline: 8455963]

52. EQ-5D-3L. EuroQol. URL: https://euroqol.org/eq-5d-instruments/eq-5d-3l-about/ [accessed 2021-02-25]

53. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [Medline: 6880820]

54. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. J Pers Assess 1985 Feb;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]

55. Thomas Antérion C, Ribas C, Honoré-Masson S, Million J, Laurent B. Evaluation de la plainte cognitive de patients Alzheimer, de sujets MCI, anxiodépressifs et de témoins avec le QPC (Questionnaire de Plainte Cognitive). Neurologie - Psychiatrie - Gériatrie 2004 Apr;4(20):30-34. [doi: 10.1016/s1627-4830(04)97931-7]

56. Bouhassira D, Attal N, Fermanian J, Alchaar H, Gautron M, Masquelier E, et al. Development and validation of the neuropathic pain symptom inventory. Pain 2004 Apr;108(3):248-257. [doi: 10.1016/j.pain.2003.12.024] [Medline: 15030944]

57. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988 Dec;15(12):1833-1840. [Medline: 3068365]

58. Gossec L, Paternotte S, Aanerud GJ, Balanescu A, Boumpas DT, Carmona L, et al. Finalisation and validation of the rheumatoid arthritis impact of disease score, a patient-derived composite measure of impact of rheumatoid arthritis: a EULAR initiative. Ann Rheum Dis 2011 Jun;70(6):935-942. [doi: 10.1136/ard.2010.142901] [Medline: 21540201]

59. Roland M, Morris R. A study of the natural history of back pain. part I: development of a reliable and sensitive measure of disability in low-back pain. Spine (Phila Pa 1976) 1983 Mar;8(2):141-144. [doi: 10.1097/00007632-198303000-00004] [Medline: 6222486]

60. Francis CY, Morris J, Whorwell PJ. The irritable bowel severity scoring system: a simple method of monitoring irritable bowel syndrome and its progress. Aliment Pharmacol Ther 1997 May;11(2):395-402 [FREE Full text] [doi: 10.1046/j.1365-2036.1997.142318000.x] [Medline: 9146781]

61. Burckhardt CS, Clark SR, Bennett RM. The fibromyalgia impact questionnaire: development and validation. J Rheumatol 1991 May;18(5):728-733. [Medline: 1865419]

62. Kosinski M, Bayliss MS, Bjorner JB, Ware JE, Garber WH, Batenhorst A, et al. A six-item short-form survey for measuring headache impact: the HIT-6. Qual Life Res 2003 Dec;12(8):963-974. [Medline: 14651415]

63. Hurst H, Bolton J. Assessing the clinical significance of change scores recorded on subjective outcome measures. J Manipulative Physiol Ther 2004 Jan;27(1):26-35. [doi: 10.1016/j.jmpt.2003.11.003] [Medline: 14739871]

64. Blanchard EB, Jones-Alexander J, Buckley TC, Forneris CA. Psychometric properties of the PTSD checklist (PCL). Behav Res Ther 1996 Aug;34(8):669-673. [doi: 10.1016/0005-7967(96)00033-2] [Medline: 8870294]

65. Bouhassira D, Attal N, Alchaar H, Boureau F, Brochet B, Bruxelle J, et al. Comparison of pain syndromes associated with nervous or somatic lesions and development of a new neuropathic pain diagnostic questionnaire (DN4). Pain 2005 Mar;114(1-2):29-36. [doi: 10.1016/j.pain.2004.12.010] [Medline: 15733628]

66. Perrot S, Bouhassira D, Fermanian J, CEDR (Cercle d'Etude de la Douleur en Rhumatologie). Development and validation of the Fibromyalgia Rapid Screening Tool (FiRST). Pain 2010 Aug;150(2):250-256. [doi: 10.1016/j.pain.2010.03.034] [Medline: 20488620]

67. Knisely JS, Wunsch MJ, Cropsey KL, Campbell ED. Prescription opioid misuse index: a brief questionnaire to assess misuse. J Subst Abuse Treat 2008 Dec;35(4):380-386. [doi: 10.1016/j.jsat.2008.02.001] [Medline: 18657935]

68. Webster LR, Webster RM. Predicting aberrant behaviors in opioid-treated patients: preliminary validation of the opioid risk tool. Pain Med 2005;6(6):432-442. [doi: 10.1111/j.1526-4637.2005.00072.x] [Medline: 16336480]

69. Thayaparan AJ, Mahdi E. The patient satisfaction questionnaire short form (PSQ-18) as an adaptable, reliable, and validated tool for use in various settings. Med Educ Online 2013 Jul 23;18:21747 [FREE Full text] [doi: 10.3402/meo.v18i0.21747] [Medline: 23883565]

70. Boß L, Lehr D, Reis D, Vis C, Riper H, Berking M, et al. Reliability and validity of assessing user satisfaction with web-based health interventions. J Med Internet Res 2016 Aug 31;18(8):e234 [FREE Full text] [doi: 10.2196/jmir.5952] [Medline: 27582341]

71. Larsen DL, Attkisson CC, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Eval Program Plann 1979;2(3):197-207. [Medline: 10245370]

72. Minen MT, Schaubhut KB, Morio K. Smartphone based behavioral therapy for pain in multiple sclerosis (MS) patients: A feasibility acceptability randomized controlled study for the treatment of comorbid migraine and ms pain. Mult Scler Relat Disord 2020 Nov;46:102489. [doi: 10.1016/j.msard.2020.102489] [Medline: 32950893]

73. Ward JH. Hierarchical grouping to optimize an objective function. J Am Stat Assoc 1963 Mar;58(301):236-244. [doi: 10.1080/01621459.1963.10500845]

74. Von Korff M, Scher AI, Helmick C, Carter-Pokras O, Dodick DW, Goulet J, et al. United States national pain strategy for population research: concepts, definitions, and pilot data. J Pain 2016 Oct;17(10):1068-1080. [doi: 10.1016/j.jpain.2016.06.009] [Medline: 27377620]

75. Shaygan M, Jaberi A. The effect of a smartphone-based pain management application on pain intensity and quality of life in adolescents with chronic pain. Sci Rep 2021 Mar 23;11(1):6588 [FREE Full text] [doi: 10.1038/s41598-021-86156-8] [Medline: 33758322]

76. Mariano TY, Wan L, Edwards RR, Jamison RN. Online teletherapy for chronic pain: a systematic review. J Telemed Telecare 2019 Sep 05:1357633X19871746. [doi: 10.1177/1357633X19871746] [Medline: 31488004]

77. Dany L, Roussel P, Laguette V, Lagouanelle-Simeoni M, Apostolidis T. Time perspective, socioeconomic status, and psychological distress in chronic pain patients. Psychol Health Med 2016;21(3):295-308. [doi: 10.1080/13548506.2015.1062900] [Medline: 26153853]

78. Frettlöh J, Maier C, Gockel H, Zenz M, Hüppe M. [Characterization of chronic pain patients in German pain centers : core data from more than 10,000 patients]. Schmerz 2009 Dec;23(6):576-591. [doi: 10.1007/s00482-009-0836-z] [Medline: 19802633]

79. May C, Brcic V, Lau B. Characteristics and complexity of chronic pain patients referred to a community-based multidisciplinary chronic pain clinic. Can J Pain 2018 Apr 19;2(1):125-134. [doi: 10.1080/24740527.2018.1453751]

80. Malon J, Shah P, Koh WY, Cattabriga G, Li E, Cao L. Characterizing the demographics of chronic pain patients in the state of Maine using the Maine all payer claims database. BMC Public Health 2018 Jun 28;18(1):810 [FREE Full text] [doi: 10.1186/s12889-018-5673-5] [Medline: 29954350]

81. Dahlhamer J, Lucas J, Zelaya C, Nahin R, Mackey S, DeBar L, et al. Prevalence of chronic pain and high-impact chronic pain among adults - United States, 2016. MMWR Morb Mortal Wkly Rep 2018 Sep 14;67(36):1001-1006 [FREE Full text] [doi: 10.15585/mmwr.mm6736a2] [Medline: 30212442]

82. Luque-Suarez A, Martinez-Calderon J, Falla D. Role of kinesiophobia on pain, disability and quality of life in people suffering from chronic musculoskeletal pain: a systematic review. Br J Sports Med 2019 May;53(9):554-559. [doi: 10.1136/bjsports-2017-098673] [Medline: 29666064]

83. Di Tella M, Castelli L. Alexithymia in chronic pain disorders. Curr Rheumatol Rep 2016 Jul;18(7):41. [doi: 10.1007/s11926-016-0592-x] [Medline: 27215759]

84. Ibrahim ME, Weber K, Courvoisier DS, Genevay S. Big five personality traits and disabling chronic low back pain: association with fear-avoidance, anxious and depressive moods. J Pain Res 2020;13:745-754 [FREE Full text] [doi: 10.2147/JPR.S237522] [Medline: 32346307]

85. Stålnacke B. Life satisfaction in patients with chronic pain - relation to pain intensity, disability, and psychological factors. Neuropsychiatr Dis Treat 2011;7:683-689 [FREE Full text] [doi: 10.2147/NDT.S25321] [Medline: 22128253]

86. Emery PC, Wilson KG, Kowal J. Major depressive disorder and sleep disturbance in patients with chronic pain. Pain Res Manag 2014;19(1):35-41 [FREE Full text] [doi: 10.1155/2014/480859] [Medline: 24367795]

87. Mutubuki EN, Beljon Y, Maas ET, Huygen FJPM, Ostelo RWJG, van Tulder MW, et al. The longitudinal relationships between pain severity and disability versus health-related quality of life and costs among chronic low back pain patients. Qual Life Res 2020 Jan;29(1):275-287 [FREE Full text] [doi: 10.1007/s11136-019-02302-w] [Medline: 31531837]

88. Baker KS, Gibson SJ, Georgiou-Karistianis N, Giummarra MJ. Relationship between self-reported cognitive difficulties, objective neuropsychological test performance and psychological distress in chronic pain. Eur J Pain 2018 Mar;22(3):601-613. [doi: 10.1002/ejp.1151] [Medline: 29160603]

89. Bargiel-Matusiewicz K, Krzyszkowska A. Dispositional optimism and coping with pain. Eur J Med Res 2009 Dec 07;14 Suppl 4:271-274 [FREE Full text] [doi: 10.1186/2047-783x-14-s4-271] [Medline: 20156771]

90. McPartland JM, Matias I, Di Marzo V, Glass M. Evolutionary origins of the endocannabinoid system. Gene 2006 Mar 29;370:64-74. [doi: 10.1016/j.gene.2005.11.004] [Medline: 16434153]

91. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]

92. Schoth DE, Radhakrishnan K, Liossi C. A systematic review with subset meta-analysis of studies exploring memory recall biases for pain-related information in adults with chronic pain. Pain Rep 2020;5(2):e816 [FREE Full text] [doi: 10.1097/PR9.0000000000000816] [Medline: 32440609]

93. Babel P. The effect of positive affect on the memory of pain. Pain Manag Nurs 2017 Jun;18(3):129-136. [doi: 10.1016/j.pmn.2017.02.198] [Medline: 28431903]

94. Jamison RN, Raymond SA, Slawsby EA, McHugo GJ, Baird JC. Pain assessment in patients with low back pain: comparison of weekly recall and momentary electronic data. J Pain 2006 Mar;7(3):192-199. [doi: 10.1016/j.jpain.2005.10.006] [Medline: 16516825]

95. Hovasapian A, Levine LJ. Reappraisal mitigates overestimation of remembered pain in anxious individuals. Cogn Emot 2016 Sep;30(6):1222-1231. [doi: 10.1080/02699931.2015.1049937] [Medline: 26192160]

96. Broderick JE, Schwartz JE, Vikingstad G, Pribbernow M, Grossman S, Stone AA. The accuracy of pain and fatigue items across different reporting periods. Pain 2008 Sep 30;139(1):146-157 [FREE Full text] [doi: 10.1016/j.pain.2008.03.024] [Medline: 18455312]

97. Gedney JJ, Logan H. Pain related recall predicts future pain report. Pain 2006 Mar;121(1-2):69-76. [doi: 10.1016/j.pain.2005.12.005] [Medline: 16480824]

98. Rode S, Salkovskis PM, Jack T. An experimental study of attention, labelling and memory in people suffering from chronic pain. Pain 2001 Nov;94(2):193-203. [doi: 10.1016/S0304-3959(01)00356-6] [Medline: 11690733]

99. Labbé É, Moulin JJ, Guéguen R, Sass C, Chatain C, Gerbaud L. Un indicateur de mesure de la précarité et de la « santé sociale » : le score EPICES: L'expérience des Centres d'examens de santé de l'Assurance maladie. La Revue de l'Ires 2007;53:3-49. [doi: 10.3917/rdli.053.0003]

100. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc 2016;9:211-217 [FREE Full text] [doi: 10.2147/JMDH.S104807] [Medline: 27217764]

101. Zhu K, McKnight B, Stergachis A, Daling JR, Levine RS. Comparison of self-report data and medical records data: results from a case-control study on prostate cancer. Int J Epidemiol 1999 Jun;28(3):409-417. [doi: 10.1093/ije/28.3.409] [Medline: 10405842]

Abbreviations

mHealth: mobile health

XSL•FO
RenderX

Original Paper

Internet-Based Information Behavior After Pregnancy Loss: Interview Study

Nazanin Andalibi[1], BSc, MSc, PhD; Kristen Bowen[2], MSLS

[1]School of Information, University of Michigan, Ann Arbor, MI, United States
[2]University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Nazanin Andalibi, BSc, MSc, PhD
School of Information
University of Michigan
105 S State St
Ann Arbor, MI, 48109
United States
Phone: 1 (734) 763 2285
Email: andalibi@umich.edu

Abstract

Background: Information behavior describes all human behaviors in relation to information. Individuals experiencing disruption or stigma often use internet-based tools and spaces to meet their associated information needs. One such context is pregnancy loss, which, although impactful and common, has been absent from much of feminist and reproductive health and information behavior scholarship. By understanding information behavior after pregnancy loss and accounting for it in designing internet-based information spaces, we can take a meaningful step toward countering the stigma and silence that many who experience such loss endure, facilitate coping, and make space for diverse pregnancy narratives in our society.

Objective: This study's objective is to provide a characterization of internet-based information behavior after pregnancy loss.

Methods: We examined internet-based information behavior after pregnancy loss through 9 in-depth interviews with individuals residing in the United States. We analyzed the data by using open and axial coding.

Results: We identified the following three themes in relation to participants' information behavior in internet-based spaces: needed information types, information-related concerns, and information outcomes. We drew from information behavior frameworks to interpret the processes and concerns described by participants as they moved from recognizing information needs to searching for information and to using information and experiencing outcomes. Specifically, we aligned these themes with information use concepts from the information behavior literature—information search, knowledge construction, information production, information application, and information effects. Participants' main concerns centered on being able to easily find information (ie, searchability), particularly on topics that had already been covered (ie, persistence), and, once found, being able to assess the information for its relevance, helpfulness, and credibility (ie, assessability). We suggest the following design implications that support health information behavior: assessability, persistence, and searchability.

Conclusions: We examined internet-based information behavior in the context of pregnancy loss, an important yet silenced reproductive health experience. Owing to the prevalence of information seeking during pregnancy, we advocate that generic pregnancy-related information spaces should address the needs related to pregnancy loss that we identified in addition to spaces dedicated to pregnancy loss. Such a shift could not only support those who use these spaces to manage pregnancies and then experience a loss but also help combat the silence and stigma associated with loss and the linear and normative narrative by which pregnancies are often represented.

(JMIR Form Res 2022;6(3):e32640)   doi:10.2196/32640

Introduction

Background

When humans face uncertainty, disruption, stigma, or distress, helpful information encounters can be beneficial in making sense of their experiences and assist in deciding how to move forward. These encounters can occur through passive (eg, finding information without seeking information) or active information seeking. Information behavior is a concept that describes all human behavior related to information, including active seeking, browsing, and information use [1]. Social media and other internet-based tools such as search engines are important to modern information seekers, especially when information needs are not met through traditional means [2,3] and when a topic may carry a stigma in a given culture and context. This is particularly true for health information behavior.

In fact, 59% of US adults report seeking health information on the internet; 39% use the internet to determine their diagnosis, and 53% of these *internet-based diagnosers* use the information found on the internet in discussions with clinicians [4]. It is well-established that individuals experiencing health conditions (which can be disruptive, distressing, and stigmatizing) use internet-based tools to assist in managing their experiences. For example, prior work at the intersection of health and information behavior provides insights into information seeking [5], social support exchange [6,7], and how to design for better search [8-10] or educational experiences [11], including the utility of internet-based spaces in a range of reproductive health contexts (eg, infertility [12,13] or pregnancy while diabetic [14]). In this work, we examine health information behavior (including seeking, sharing, using, and the impact) as mediated and facilitated through internet-based spaces and tools in a key reproductive health context—pregnancy loss—for which such behavior remains to be explored.

Pregnancy loss is often colloquially referred to as *miscarriage.* By *pregnancy loss,* we mean the unintended loss of a desired pregnancy at any gestational stage. Pregnancy loss is an important context for investigating information behavior as it is distressing, stigmatizing, disruptive, and traumatizing for many. It challenges people's identities as expecting parents and changes interpersonal relationships but is generally absent from societal narratives of grief or reproduction [15]. Furthermore, individuals experiencing pregnancy loss report negative interactions with medical providers, family, and friends [15]. We aligned our work with that of Layne [16], a feminist anthropologist, and Andalibi [17], a human-computer interaction (HCI) researcher, who argued that pregnancy loss should be included in feminist reproductive health discourse and pregnancy-related technology design, respectively. Indeed, the analysis of pregnancy-related mobile apps by Andalibi [17] illustrates the lack of consideration for pregnancy loss as a possible outcome of pregnancy in most apps that individuals with pregnancies use and argues that this reinforces linear, normative narratives about pregnancy and is marginalizing and harmful. The sociocultural context surrounding pregnancy loss leaves many without basic information about it, which can complicate and hinder coping.

Prior information behavior research concerning stigmatizing and distressing experiences [2,18,19] has explored the challenges people face when engaging with authority figures (eg, care providers) and how they can leverage internet-based information to fill needs not met in clinical encounters. Such work provides insights into the information that specific groups seek [20,21], how privileged and other communities differ in their use of information to inform their care [22], and the benefits and challenges individuals face when finding information on the internet [12,21,23,24]. With respect to pregnancy loss, prior work has examined topics such as social support and self-disclosure at the intersection of pregnancy loss and technology [25-27], showing how internet-based spaces are sites for support exchange, sensemaking, and validation or invalidation [28-30]. A recent study provided insights into information needs after loss, including general information about loss, counseling resources, others' experiences, and information from providers [28]. We built on these works to explore the dimensions of information behavior after pregnancy loss in relation to existing internet-based spaces, the needs these spaces help meet, challenges they pose to individuals enduring a pregnancy loss, and outcomes.

We conducted interviews with women in the United States who had recently experienced a pregnancy loss and who used social media. We reported on 3 themes related to needed information types, information-related concerns, and information outcomes as mediated through internet-based spaces and tools. We interpreted these themes by drawing from conceptions of information use synthesized by Kari [31]: information search, knowledge construction, information production, applying information, and the effects of information. We found two main categories of information need—evidence-based and experience-based information—echoing health information needs research in other contexts [32,33]. Participants' main information concerns were related to their ability to find information and their ability to assess the credibility of the information they found. Finally, we found that internet-based information encounters could lead to not only finding answers to questions but also learning what questions to ask and sometimes how to advocate for oneself in clinical encounters or even consider advocating for oneself as an option.

By examining information behavior in relation to experiencing personal uncertainty, disruption, stigma, or distress, we can better understand the ways people go about meeting their information needs, thus uncovering the challenges and benefits they face in day-to-day information encounters. This knowledge will (1) allow technologists and researchers to design technologies that meet people's information needs in times of distress and fit into their existing practices with the potential to positively affect their well-being and (2) improve clinicians' understanding of the pregnancy loss experience.

It should be noted that we advocate that pregnancy loss is not just a women's health topic—individuals not identifying as women can also experience pregnancy and loss, and these experiences are important. In this study, although our recruitment efforts were not limited to individuals identifying as women, all participants were women. Therefore, we used this framing and terminology throughout this paper.

Related Work

Information Behavior

Wilson [1] presented a general model of information behavior comprising information need, seeking, processing, and use and consideration of the role that an individual's environment can play in their information behavior. The model by Wilson [1] has become a foundation for researchers exploring information behavior. However, Wilson [1] only accounts for when one has a known information need. Erdelez [34] extended the information behavior theories to encompass the processing of information that one was not actively seeking, termed *information encountering*. This concept allows for the exploration of the use of technologies in an information-saturated environment [34-36].

Prior scholarships have examined information behavior by examining information use. Kari [31] conducted a systematic literature review on information use within the information studies field and outlined 7 major conceptions that we draw from in interpreting our findings. In this paper, we explore the health information behavior of individuals who have experienced pregnancy loss and seek information on the internet. These conceptions, proposed by Kari [31], are relevant to our study of information behavior related to pregnancy loss:

- Information search: the processes of information seeking and information retrieval
- Knowledge construction: mental constructs are shaped or designed to function as a basis of thinking
- Information production: creating an expression of knowledge, which others can also observe
- Applying information: information functions as a resource in some processes
- Effects of information: changes brought about by information

HCI scholarship's examination of information behavior has mainly centered on the concepts of information need [28] and information search (eg, how social contexts, environments, specific disabilities, or even mood can influence individuals' information-seeking behavior [8-10,37,38]). In this study, we recognized the need to go beyond information seeking and consider information use concepts.

Health Information Behavior and Internet-Based Spaces

Health information behavior encompasses the general model created by Wilson [1] but with the information component being health information, which includes how individuals seek information about their health and health risks and examines their engagement or disengagement with health information in connection to information use concepts [31,39]. Health information seekers, both offline and on the internet, admit a heavy reliance on medical professionals for information. Medical professionals are often the first choice; however, the prevalence of internet-based resources and an optional social component has made internet-based health information seeking widespread in the United States [40,41]. Marginalized and stigmatized communities may especially benefit from finding additional health information, which they can present to their medical professionals to counteract language barriers, bias, and

general dismissal [20,42]. Therefore, these groups are primed to become innovative information seekers, as they are likely to turn to and trust outside sources of health information [20-22]. Internet-based spaces can provide sources for information and support for those who are denied it or who do not sufficiently trust sources such as medical professionals to consult them [27,43,44].

Prior research has identified two types of health information: expertise-based information, which is produced by medical professionals, and experience-based information, which refers to information shared based on subjective first-hand health experiences [32]. People use both evidence-based health information and experience-based health information found on internet sites, social networking sites, and blogs [33]. Experience-based information also affects people's health behaviors [33]. Examining health information behavior may allow us to better design health information needs.

Previous scholarship has emphasized investigating the behavior of medical professionals to better design future technology that can assist in streamlining their processes [45-48]. Noted as early as 10 years before this study [49], investigating the context of health information searching and use by individuals experiencing health conditions is also necessary to design information spaces on the internet in support of health. There have been developments in research on consumer health information behavior, even if not framed using information behavior theories, with communities such as older adults [50], pregnancy (excluding loss) [51,52], and HIV information seekers [24]. There remains a need to examine the breadth of health information topics, depth of health information seeker communities, and varied behaviors and outcomes once information is discovered. A context that is in and of itself important but that can also teach us about other stigmatizing and marginalized health experiences is pregnancy loss.

Pregnancy, Pregnancy Loss, and Information Behavior in Internet-Based Spaces

Information behavior scholarship related to coping with pregnancy loss is scarce, even as scholars examine the information behavior of individuals with pregnancies. Women with pregnancies use internet resources to manage pregnancies [53,54]. In fact, social media and other internet-based resources play an important role in the information needs of expectant individuals in the United States [53-56].

The information needs of individuals with pregnancies have been a research focus [57,58], although only with generic mentions of pregnancy loss, with few exceptions [28]. Similarly, design research [59,60] in the pregnancy space has largely not accounted for pregnancy loss. The significance of pregnancy loss is often minimized and rendered *invisible*. Gold et al [61] were among the first to examine internet-based support seeking of individuals who have experienced pregnancy loss, finding that internet-based support can provide a safe haven from in-person stigma and revealing a preference for moderators within these spaces. This highlights the need to understand pregnancy loss–related information needs and behavior, especially as there is no unique way in which loss is experienced [28].

We focus on pregnancy loss to examine how individuals who are enduring a loss use internet-based spaces to meet their information needs and identify the challenges and outcomes they face in this process. We effectively examine information behavior mediated through internet-based spaces for individuals enduring a pregnancy loss. Pregnancy loss, in this sense, is more broadly situated within both reproductive health and women's health.

Pregnancy loss is common, occurring in approximately 20% of the identified pregnancies [62]. It is linked to stigma, mental health challenges, and shifts in identity and personal relationships [15]. It is also difficult to find support because of the stigma, shame, and guilt attached to it, although it tends to be traumatizing [63]. Not only is the experience absent from dominant pregnancy narratives [16], but also grieving such losses is not socially acceptable, leading to inconsiderate or unsupportive reactions from others [15]. Research at the intersection of pregnancy loss and technology has examined how and why people choose to disclose it on social media [26,27,64], how and why others respond to such content [65], the outcomes of such disclosures [25], pregnancy-related apps' inclusion of pregnancy loss [17], validation seeking on the internet after a loss [30], and information needs [28]. Kresnye et al [28] found that forums, Facebook, and blogs are common internet-based sources after pregnancy loss and that people needed general loss-related information, counseling resources, information about others' experiences, and information from medical providers. However, they also found challenges such as difficulty in locating resources and stigma. We have built on such work to examine what needs are met by existing internet-based spaces and what challenges they provide. We address the following research question: What are information-related processes (eg, needs, challenges, and outcomes) for individuals enduring pregnancy losses, and how are these mediated through internet-based tools?

Methods

Recruitment

We conducted 9 remote in-depth semistructured interviews with individuals who had experienced pregnancy loss or losses within the past 2 years (to facilitate recall and similar social media landscape). Interviews averaged 92 (range: 80-104) minutes. In this study, pregnancy loss was defined as the unintentional loss of a pregnancy at any gestational stage. Research has suggested that the gestational stage of loss is not a causal factor in the severity of grief and coping experience [66]; therefore, we did not screen participants based on the loss stage. We used a screening survey to purposefully [67] sample participants who varied in experience, demographics, and technology use to the extent possible. The screening survey was shared on the lead researcher's social media accounts and, from there, beyond their network. This choice was informed by prior work [68-70] that used social media for reaching hard to reach populations. This was also appropriate as a criterion for inclusion was social media use. We do not know how many times the link was shared or who shared it. We did not ask participants to share the study call with those who may be eligible to participate.

Ethics Approval

This study was approved by the University of Michigan Institutional Review Board (HUM00156077).

Screening Survey

To qualify, respondents had to meet the following requirements: live in the United States, have experienced a pregnancy loss within the past 2 years, be a social media user, and be aged at least 18 years. Only respondents who met these criteria went on to the next stage of the screening survey, which included questions about the general use of internet-based platforms, internet-based platforms used in connection to pregnancy loss, needs after a loss to assist in processing, the month and year of the most recent pregnancy loss, age, self-description of gender, race, ethnicity, status within the LGBTQ (lesbian, gay, bisexual, transgender, and queer or questioning) community, self-description of whether they had children, relationship status, education level, household income, primary religion, and residence in an urban area (>50,000 residents) or rural areas (<50,000). We asked respondents to self-describe whether they had children or not (and how many) as this can be an emotionally charged question for those who have experienced pregnancy losses and to allow for diverse ideological views on the matter.

The screening survey received 49 responses. After survey completion, the lead researcher contacted respondents with an internet-based consent form, additional study information (eg, tools and devices they would need access to during the interview), and scheduling options. A total of 9 people completed this process and were included in the study. Participant ages ranged from 31 to 42 years. Of the 9 individuals, 8 identified as Caucasian or White and 1 as Black American; 5 identified as having children, 2 stated that they had none, and 2 chose not to answer. All identified as women, and all had some college education, with most having an advanced degree. Most were married, and one of the individuals identified as single. Of the 9 individuals, 1 identified as LGBTQ and 8 as heterosexual and cisgender. Of the 9 individuals, 1 made <US $50,000; the rest had an income of ≥US $75,000. Most lived in urban areas, and one of the individuals lived in a rural area. Table 1 presents the participants' information.

Table 1. Participant information.

Characteristics	Participant ID								
	P1	P2	P3	P4	P5	P6	P7	P8	P9
Age (years)	31	34	42	32	37	30	39	31	37
Gender[a]	Female	Female	Female	Female	Female	Female	Female	Female	Female
Race[a]	Caucasian	Caucasian	White	Caucasian	Caucasian	Black	White	White	White
Ethnicity[a]	White	Adopted	Ashkenazi Jewish	Jewish	White	Black American	White	White	White
LGBTQ[b]	No	No	No	No	No	Yes	No	No	No
Children[a]	No answer	2	1	0	1	No answer	1	1	0
Relationship status	Married	Married	Married	Married	Single	Married	Married	Married	Married
Education	Graduate degree	Graduate degree	College	College	Some college	Graduate degree	Graduate degree	Graduate degree	College
Income (US $)	≥75,000	≥75,000	≥75,000	≥75,000	30,000-49,999	≥75,000	≥75,000	≥75,000	≥75,000
Religion	Christian	Catholic	Jewish	Jewish	Catholic	Atheist	Agnostic	Nothing	Atheist
Rural or urban	Urban	Urban	Urban	Rural	Urban	Urban	Urban	Urban	Urban
Technology uses	FB[c], FSG[d], IG[e], and fertility tracking apps	FB, FSG, IG, and fertility or pregnancy apps	FB, FSG, IG, and TW[f]	FB, IG, Snapchat, pregnancy apps, and forums	FB, FSG, IG, TW, and fertility or pregnancy apps	FB, IG, TW, and pregnancy apps	FB, FSG, IG, TW, fertility apps, and forums	FB, IG, and infertility or pregnancy apps	FB, FSG, IG, pregnancy apps, and forums

[a]Responses to questions about gender, race, ethnicity, and having children were open ended to provide flexibility for self-descriptions. For others, we used predefined choices with an option to self-describe.

[b]LGBTQ: lesbian, gay, bisexual, transgender, and queer or questioning.

[c]FB: Facebook.

[d]FSG: Facebook support groups.

[e]IG: Instagram.

[f]TW: Twitter.

Interview Overview

Participants chose their preferred remote tool (eg, Skype or Google Meet) as long as it allowed viewing our screen when shared (required for a portion of the study that was not used in this paper). To preserve participants' privacy and facilitate their comfort, only audio was recorded.

Interview Guide Overview

First, participants were asked about their life when they found out they were pregnant and what followed. Next, participants were invited to 3 illustration tasks, with their choice of words or drawing on paper, their general social media use, and their technology use during pregnancy and in relation to pregnancy loss, respectively. After each illustration, participants were asked to photograph their work and send it (via email and text) to the interviewer during the call. They were then asked to provide detailed verbal descriptions of their submissions. Visual elicitation methods often help participants articulate mental models associated with sensitive topics, facilitating flexibility and reflection [71-73]. Studies on pregnancy support networks [74] have used similar methods. We emphasized that only *they*

needed to understand their submission and that we would ask them to explain it to us later.

Next, we focused on understanding participants' needs for coping and support after loss and how these needs were or were not met, including through the use of technology. Finally, participants used paper and pens to outline an ideal support system and then shared and explained these to us, as in the aforementioned activities. Here, we focus only on the results of information behavior.

Analysis

We recorded the audio during interviews and transcribed them for analysis. We conducted a qualitative analysis using open coding, followed by axial coding [75]. That is, we did not base our analysis on pre-existing codes as it was important to center participants' voices and pregnancy loss.

Guided by our research question, the first author (who also conducted the interviews) developed initial codes by open coding 3 interviews. They iterated these codes and refined them using memoing in the process. Then, a team member trained in qualitative analysis was given the data and codes who met with the first author to discuss the definitions. The team member

spent several weeks familiarizing themselves with all data and codes, discussing them weekly with the first author. After gaining a reasonable understanding, the team member independently applied the codes to the same 3 interviews coded by the first author. They created no new codes, despite keeping an eye out for new observations. The first author and the team member discussed any divergence in the coding of these 3 interviews to ensure mutual understanding. Then, the team member coded the rest of the data, discussing the interviews and code applications with the first author weekly. The first author then assessed the codes and excerpts to ensure consistency and grouped codes into larger categories. For example, codes for various information needs were grouped under the umbrella of information needs. The authors then refined the observed themes and discussed them in relation to the Kari [31] framework; in other words, we did not set out to use the Kari [31] framework in our study as we did not know what would be important to participants; we collected the data, analyzed them, and consulted the work by Kari [31] to interpret and organize our resulting themes. In this paper, we report themes related to our research question, focusing on information behavior.

Results

Overview

We report on our findings in three main themes related to the participants' information behavior: (1) needed information type, (2) information-related concerns, and (3) information outcomes. In interpreting these themes, we draw connections to the five information use conceptions conceived by Kari [31], presented in the *Information Behavior* section: knowledge construction, information search, applying information, information production, and effects of information. We describe how these concepts illuminate participants' use of internet-based spaces (ie, forums, pregnancy apps, search engines, and Facebook groups) for information related to pregnancy loss and coping, grieving, and sense making.

Knowledge Construction and Needed Information Types

Overview

An unmet need to gain knowledge motivated participants to turn to internet-based spaces to seek information. Individuals who are likely to face discrimination or stigmatization in health care settings often turn to other sources of information [2,3]. Our findings suggest that the main types of information participants sought fell within these categories: (1) science and evidence-based information about pregnancy loss, including information about the individual with pregnancy and their partner, and (2) experience-based information, including information about pregnancy-related medical conditions and others' experiences with pregnancy loss. According to the concept of *knowledge construction* by Kari [31], individuals turn to information seeking (ie, information search) to resolve unknown or unsure knowledge and enter the process of developing new knowledge, ideas, and beliefs.

Information sources for these needs were others' stories and experiences (experience-based) and scientifically oriented information (evidence-based), the categories in which knowledge construction [31] occurred. Notably, P9 said, "First and foremost I would say, I needed...information and...to hear other people's stories."

Similarly, 2 broad types of information paths were identified; P8 noted the following:

> So then there's the two paths...on the one side was that the evidence based in...things that were actually true like journal article or something that had qualitative and quantitative research...and that, in some ways, that led to this down arrow of information, which...was helpful...at times. But then after you become a statistic like losing two...sometimes the research doesn't mean as much anymore because you realize that you can be a two percent....

P8 continued by stating the following:

> But then on the other side was the real experiences and that's what I would find through some of those forums...that in some ways it didn't lead to that evidence-based information, but it led to in some ways validation.

Other people and academic, evidence-based information were important for P8's experience and were complementary. In fact, many differentiated between medical, evidence-based information, and other types. For example, P7 noted a distinction between the 2 information sources:

> And I think...the practical information doesn't necessarily have to be medically informed. Having towels under you when you sleep doesn't need to be medically informed.

This speaks to information that one would need that is not necessarily sourced from a medical professional but is still key. In what follows, we delve deeper into the types of information that participants sought on the internet.

Science- and Evidence-Based Information Needs

Participants often felt that medical professionals did not adequately meet their information needs:

> I think, if it were to happen to me again, I would wanna talk to my OB at length, like, "Why is it that some people do it, some people don't?"...stuff like that, being able to know specifically related to medical procedures, under what circumstances should one have a D&C? What are the implications of it? 'Cause I've heard that there's some negative side effects. [P7]

P7 had questions not answered by her provider about dilation and curettage. She turned to internet-based spaces such as support groups, forums, and search engines.

Medical information needs were also informed by the gap between the participants' needs and their level of trust in their medical care provider. Combined with the relationship aspect is a hesitancy and inability to reach out to their medical care

provider when concerns may be viewed as trivial but are nonetheless crucial to participants' experiences. Seeking evidence-based information via Google and other internet-based resources such as online support groups and forums was a way of complementing the information received from medical professionals.

P1 spoke of the role that her provider could have played to meet her information needs but did not:

> *I think that would have done wonders. Having my doctor able to answer all the little insane questions I would have all the time because when you're feeling things that you're not used to feeling...I automatically thought something was wrong...Like oh, this is happening to my body, let me Google it because I can't just call my doctor every five minutes. So, having access to a healthcare provider that doesn't mind you badgering them.*

Obtaining knowledge about loss before both pregnancy and loss was also important. A lack of such information increased confusion and difficulty when the loss occurred. As we see in P5's remarks, being kept in the dark by medical providers about the risks for and signs of pregnancy loss prompted her to seek information and others' stories via Google:

> *When I went into the first pregnancy, I was very oblivious to a lot of stuff...I wish I would have had the knowledge, either from the group or by reading, researching, looking things up...I just assumed everything's great, you're pregnant...my doctor pretty much kept me in the dark about it. He would just, "Okay. Everything's fine. Everything's fine." When, in reality, it wasn't. I had to learn that the second time around from other people.*

For her first pregnancy, P5 trusted that her health care provider had shared the needed knowledge and that trust contributed to P5 not looking for outside information. However, after experiencing the first pregnancy loss, P5 blamed herself for not independently seeking evidence-based information. P5 had unmet medical information needs, which were only learned after her experience with pregnancy loss. P5 expressed how much she would have benefited from the missing knowledge in coping with the first pregnancy loss. These accounts align with the finding by Kresnye et al [28] that individuals tend to attempt to find information after a loss occurs, not before it.

Some participants also noted an unmet need for information from their partners. This could take the form of resources for conveying the information they received from a medical provider to their partners and specific signs a partner could look out for.

P6 said the following:

> *I think, also information to help the partners. Cause like, I would go to my husband...but he didn't have any more information than I did. And...we're both highly educated...But...all the information we had, didn't really prep us for...understanding miscarriage.*

P2 discovered that internet-based groups could help fill that need:

> *It was also nice for my husband because...I don't do a very good job of explaining my physical. I tend to write everything off. I'm like "I'm fine."...And so it was nice for him to be able to read what I was physically going through since I was just saying I was fine. And it was helpful because then when he did read it, he would say "Stop doing the dishes. Go sit down." "Don't make dinner, I'm going to pick something up tonight."*

These quotes showed that P6 would appreciate guidance from medical professionals for partners of individuals navigating pregnancy loss, whereas P2 found guidance on how to express the toll of pregnancy loss to her husband and gain the support she needed from him. While some participants viewed that they experienced the loss *together* with their partner, the partner could not always be helpful partially because of lack of information. Partners can be caregivers and supporters in pregnancy journeys [76], and we suggest that future work should further consider partners experiencing losses as part of those journeys.

The turn to internet-based spaces for knowledge construction often stemmed from the unavailability of information from health care providers, whether because of participants' hesitancy to trust health care practitioners or the health care practitioners' unwillingness to provide it. What Kari [31] refers to as knowledge construction is the processing of information to develop new knowledge or validate unsure knowledge. Science- and evidence-based information related to pregnancy loss was an unmet need high in priority to participants, but so was practical information from others' experiences. In the next section, we address the importance of the latter information type in internet-based spaces and the comfort offered by such knowledge.

Experience-Based Information Needs

A need for medically related information, which may be considered more reliable when the provider has first-hand experience or shared cultural views and identities, related to specific medical conditions or physical experiences connected to pregnancy loss. This information, sought in spaces such as blogs or internet-based support groups, was helpful, as these concerns may have been thought of as trivial by health care providers.

P3 described a need that may not easily be met by medical professionals:

> *But I feel like a lot of my issues...are kind of cultural...and that being able to speak about them in a group that's more culturally similar...like Jewish rituals, superstitions...being able to have other people's input and guidance and what worked for them would have been nice.*

P3 wanted to find information sources who understood her cultural beliefs and had insight into dealing with pregnancy loss as a Jewish person. Determining whether a medical professional

could provide this insight may not be possible, so she turned to digital spaces.

Similarly, P6 emphasized the importance of receiving information from individuals who share similarities, particularly when one is a part of a group that can often receive less than the standard of care from medical professionals.

> *Well...there's been a series of studies...about how, oh black women are nearly four times more likely to die giving birth...more likely to be denied pain medication...their babies are more likely to die before the age of one. So that's something I think about as a researcher. Like the gap between logistics, because all of those are just the outcomes of logistic models...you have to understand that statistics are not a good way to tell the story. So, I started just paying attention more to women, black women telling their stories. And...anecdotes aren't data. But those stories do matter....* [P6]

P3 and P6 turned to internet-based spaces in search of information that they deemed critical to their specific experiences of pregnancy loss based on culture, race, medical beliefs, and medical statistics. Individuals who have experienced pregnancy loss may not have access to medical professionals who can provide information that speak specifically to such cultural or racial issues.

Often, medical information could originate from social ties with relevant personal experiences (as described above) or professional expertise. For example, P2 mentioned needing dietary information after her loss, which a family member who had a relevant scientific background was able to provide:

> *My kid's godfather...studies gut bacteria and he was like "Well here are some things that can actually help with...absorption of nutrients" And so that was really helpful. If I'm taking all these supplements, then I want to know how I'm actually absorbing them.*

These examples illustrate participants' needs to find information not only about pregnancy loss broadly but also about the variety of relevant conditions that may not be shared across all individuals experiencing pregnancy losses.

Participants also reflected on the need to know the basics of pregnancy loss symptoms, what to physically expect, and how to practically prepare:

> *I wish a girlfriend would have been like, hey, you should lay down a bunch of towels in the bed because the second day is worse...it would have been very useful...But people don't tell you this...Little tiny practical tips...Nobody would hurt from...sleeping on towels.* [P7]

In addition to information aligned with one's experience, condition, and identity and culture, participants noted that information from others who had experienced losses could also be practical.

Another example included not only looking for symptoms of loss but also gaining basic information on the stages of pregnancy:

> *Basically Glow was not only a way for you to track your symptoms of your cycle...Do you have your cervical discharge?, and any kind of other symptoms that you have and if you've gotten a positive OPK. The forums are also helpful because I didn't know a lot at the time. I didn't know that it takes 6-12 days for an egg to implant and 2-3 days...to get a positive after that.* [P4]

This basic information on the symptoms of loss and pregnancy stages may be overlooked by medical professionals as obvious or trivial, which makes internet-based groups an avenue for education. Furthermore, we see that some information needs after a loss tend to be about future pregnancies, which are necessarily positioned within one's pregnancy history, including the loss.

A combination of evidence-based, identity and culture–aligned, and practical information from others who have experience with pregnancy loss made up most of the information that participants sought or needed. However, participants also noted finding information that they would have preferred from their evidence-based sources that were not readily available. For example, some were able to find information on forums that were not available through Google searches (ie, deleted by clinic businesses) and may not be readily shared by health care professionals. For instance, P7 shared the following:

> *Those forums were super-duper useful. And...that information wouldn't be out there at Google, because these fertility clinics...don't want people to actually know how much these things cost. And...They clean the web of reviews and stuff, because it's a business.*

In this section, we described the information needs that health care providers were either unwilling or unable to meet and the sources that participants, in turn, used to construct knowledge. Our analysis highlights two main overarching types of information that participants sought: (1) science- and evidence-based information about pregnancy loss, including information about the pregnant individual's partner, and (2) experience-based information, including information about pregnancy-related medical conditions and others' experiences with pregnancy loss, especially those aligned with one's culture and identity. Participants were able to address many of these information needs in internet-based spaces but would appreciate and benefit from health care providers' input regarding the credibility of medically related information and their assistance in conveying their knowledge to partners. The next section covers specific concerns that participants expressed related to pregnancy loss–related information encounters in internet-based spaces.

Information-Related Concerns in Information Search

Overview

Our findings suggested two main concerns in information encounters: (1) being able to find information and (2) determining the credibility of information. These are aligned with the *information search* concept of Kari [31], as the ability to use information or information resources is directly connected to the feasibility of searching for and accessing information,

which makes these concerns important for designing internet-based spaces. Persistence (ie, "the extent to which a platform affords the continued availability of content over time" [77]) and searchability (ie, the extent to which a platform affords users the ability to search for content using search terms) together afforded timely access to information.

Challenge 1: Ability to Find Information Shaped by Persistence and Searchability

Persistence and searchability often went hand in hand. Easy access to and almost real-time return on information was important. Posting a question on an online forum or support group and waiting for responses was not an appropriate method to meet some needs.

For example, P7 used search engines to quickly find information from old forum posts:

> *For me at that stage, googling for info went infinitely to the top...I didn't necessarily wanna talk to a human about what I was experiencing...I didn't wanna wait for a human to respond. I wanted in that minute to be like, "How much blood loss is too much? When should I call my doctor?" I need this immediately.*

Participants appreciated the ability to search for information on the internet, find it quickly, and be able to access it without posting about their own experiences, which can be difficult during moments of distress. The benefit of finding and reading available posts and answers rather than posting oneself is seen in P1's response:

> *I had lots of questions. The cool thing about the Facebook group is not only do you have everything that people are posting right now, but you can go to the top and you can type in a specific thing. And then all the posts that have to do with that come up.*

P7 described the importance of the interfaces in supporting search, evidence-based information, and archived information:

> *Medically informed information with citations...in terms of the actual interface it has to be super-duper searchable, with a very good search. That's one of the downsides to app-based stuff is that you can't go into the Ovia pregnancy app for example, and search 12 weeks or something...Also...making sure that it's archived, so that...I can see here's something someone posted three years ago that was similar to my situation, so I don't have to wait for answers in the moment...there was quite a while that the searchability of Facebook was really poor, especially on mobile...that's really unfortunate.*

Participants found value in archived information, collective experiences, and searchability. Being able to search for relevant information required an effective search functionality and permanent storage that allowed participants to draw from the vast amount of experience-based and evidence-based information shared over time. Internet-based spaces can provide swift access, in contrast to contacting, scheduling an appointment, and then conversing with health care providers. If an internet-based space maintains truly usable search

functionality and stores information in a persistent manner, it can be a resource for information needed at particular moments. As seen in P7's comment on the difficulties of searching in apps, when persistence and searchability are not maintained, information seekers turn to other spaces to meet their information needs.

Challenge 2: Determining Information Credibility

Facing misinformation or information perceived as less credible or difficult to assess was another challenge. Participants often turned to internet-based forums for information about specific conditions related to loss. There, they encountered the challenge of assessing what information was credible and what was not.

P1 described this problem as follows:

> *As far as Baby Center, I used that a ton right after the ectopic pregnancy because I didn't know who else to ask...I found a lot of helpful things...but...you also find things like things that don't seem to be helpful...oh, I heard my neighbor's friend's sister had this response. Like well, is it that really...fact?*

As another example, P4 said, "[That was] really helpful in getting educated, but there's also a lot of garbage in there, which I found out once I moved over to the Ava group." Although P4 was able to gain some useful information in a Facebook group, she also found misinformation, which became apparent to her once she moved to a space that she thought included less misinformation.

The amount of misinformation led some to leave internet-based forums altogether: "I did look at the forum but they drove me nuts and I stopped looking at them. They're a lot of misinformation..." [P6]. This is problematic as people come to these spaces to find information and social support; leaving means that they will not be able to access other types of support either. Ideally, people who join these spaces would stay for some time, both to find support and information for themselves and contribute to support others.

Participants found it difficult to find medically approved information about pregnancy loss in general pregnancy forums. This was particularly challenging and a cause for confusion because of advice that was contested. To this point, P7 said the following:

> *But medically informed information that...I could trust...is hard to find in general forums. Those infertility forums are pretty exceptional in how knowledgeable those people are about medical stuff, but in general pregnancy forums, people are dumb as rocks.*

When participants sought information on Google or forums, they noticed a prevalence of what they considered nonscientific information. P3 reflected on a way of assessing the credibility of evidence-based information shared in digital spaces:

> *I could imagine, in a way, an information center, where it's not so much about feelings but the questions are more fact-based and the discussion is more fact-based. I guess you can't really have medical advice because it's always lawsuits. But where people*

are encouraged to give answers that are fact-based, like studies and what they read.

Finding credible information that one can trust could have helped P7 and P3 make sense of their experiences, plan for the next steps, and regain control that was lost through the loss experience.

A main motivator for turning to internet-based spaces was a need to gain knowledge quickly; however, that benefit can be negated when one is unable to determine the credibility of information, when the information is deemed to be not credible, or when digital spaces do not afford effective searching. In the next section, we address the outcomes of looking for and finding information on pregnancy loss in internet-based spaces.

Information Outcomes: Applying Information, Information Production, and Effects of Information

Overview

Accessing information in internet-based spaces has implications for participants' future behaviors. We found that after new knowledge construction, participants (1) used information to advocate for themselves with medical professionals and (2) had their concerns and ideas validated. These outcomes align with the following information use concepts identified by Kari [31]: applying information, which occurs when one uses new knowledge, such as raising concerns to a medical professional; information production, which involves using learned knowledge to produce information for others, such as writing a post on an internet-based space; and the effects of information, where the information influences one's future choices. In these internet-based information spaces, participants developed tools for clinical encounters, such as learning what to ask and how to advocate for themselves. Newly obtained knowledge can also provide insights into options that are not presented or even discouraged by health care providers, in turn influencing one's decisions.

Applying Information: Learning What to Ask

For some, internet-based spaces were places where participants learned what questions to ask. P9 said the following:

It's only with finding an online community that I also found the language I needed to ask, should we be doing the following things? Should we be considering this kind of doctor? This kind of testing?

This is important, as the stigma surrounding loss contributes to less discourse and education about the topic [26,63].

Learning without asking was also a benefit, described by P5 as follows:

At the beginning of my pregnancy, there was a lot of warning signs that I wasn't aware of: the heartbeat was low, I had low progesterone. There's so many different things that I didn't pick up on, because I just wasn't aware. And being part of this group...helped me become aware, because these women had similar stories.

This is noteworthy, as lurking is often associated with negative connotations [78]; however, we suggest that it can serve as an effective learning strategy.

Overall, internet-based spaces and tools not only provided answers but also inspired participants with questions they could ask health care providers. Knowing what to ask and what to look for can make an individual more informed and lead to changed behavior. Participants expressed that they had used or would use knowledge found in digital spaces to obtain the needed information in conversations with medical professionals and prepare an outline of warning signs to try to avoid a future pregnancy loss; these examples align with the Kari [31] concept of applying information.

Information Production and Effects of Information: Changed Behavior

Internet-based spaces also become spaces for sharing one's own experience to inform others. P3 used an internet-based space to share information that she thought was needed and unavailable from other sources:

The other thing I'll post about is with my third miscarriage, I had the choice of getting...a different procedure...that most women...don't realize is an option...So that's kind of my soapbox...I'll often post about it...if I feel like I can offer relevant, helpful information.

In this case, P3 produced information to help others advocate for themselves by contributing experience-based information that she knew would otherwise be unavailable to others. This is an example of the information production concept of information use by Kari [31].

Another example is found in P2's choice to remain in an infertility group as a source of information for others:

I debated...whether or not to stay, but I feel like if anyone was going through what I went through, it took so long for my doctors to diagnose the problem that...I'd want to help somebody, because it was a simple blood test that nobody bothered to run for a year.

P2's role in the infertility group changed from information seeker to information producer, a source of experience-based information.

As noted in earlier sections, internet-based spaces such as Facebook groups or forums fill an informational gap that can occur when health care professionals are not available or refuse to meet information needs and when other internet-based resources (eg, Google) are unavailable or inaccessible (eg, information related to fertility clinics deleted from Google by clinic businesses). Here, we see how the knowledge gained in these spaces can lead to more informed decisions. As an example, P9 shared a story about finding instructions about in vitro fertilization injections, which led to behavior change; this content comprised questions that she did not get answers to from her provider but instead from other patients, including how to advocate for herself with her provider:

> *But in that group...especially once we started IVF...people share tips about how to make injections less painful, how to get the right angle...I know it can sound...ill-advised to get medical advice from people who aren't doctors, but a lot of the advice is more about how to advocate for yourself as a patient...something that I have learned almost exclusively through being in these groups.*

P5 offered another example of learning to advocate for oneself:

> *I had to be put on progesterone and I had to advocate for myself with my doctor...to put me on whatever progesterone they could...to maintain this pregnancy, which I carried to full term...That was, again, something I learned from the women in those groups...You definitely want to be your own advocate, and self-advocate for yourself...*

The above examples align with the concept of *effects of information* by Kari [31], where the information participants found in these internet-based spaces influenced their future choices, including how to navigate advocacy with a physician.

Discussion

Principal Findings

Our findings highlighted the following three novel themes associated with information behavior after pregnancy loss as mediated through internet-based spaces: (1) needed information type, (2) information-related concerns, and (3) information outcomes. These findings are significant as a first step in designing internet-based spaces to account for pregnancy loss information behavior. We found that participants shrank the knowledge gap regarding what to ask by learning what information others found worth knowing. They bridged language barriers by using information from internet-based spaces to supplement suggestions from medical providers, and the barriers of social stigma, cultural taboo, and lack of social and economic capital were to some extent offset by participants' ability to advocate for themselves, stemming from the confidence gained from interactions in internet-based spaces.

We showed how internet-based information spaces help alleviate some information gaps for individuals experiencing pregnancy losses and supplement information gained in clinical encounters. Nevertheless, digital health resources pose their own obstacles: people need to be able to find the information, understand and determine its credibility, and know how to use it. In addition, primary barriers to health care for women from stigmatized groups include knowledge gaps such as not knowing where to find information and what resources are available or what information to ask for, and a lack of understanding between patients and providers [79], which we also saw examples of in the case of pregnancy loss.

Information Behavior, Health, Pregnancy, and Pregnancy Loss on the Internet

In our dedication to fully engage with information theories beyond search and need, we outlined our analysis in connection with the information use concepts by Kari [31]; however, our

participants' responses also related to other information models. The Wilson [1] model remains relevant, as seen in participant responses first in their expression of unmet information needs (ie, information types), which led to information seeking on the internet, and in their insight into how the environment in which they sought information can influence their behavior (ie, information-related concerns), and finally in how they used information (ie, information outcomes). However, participants often learned what questions they *should* ask related to pregnancy loss in these internet-based spaces, which is a form of information encountering [34]. The internet, as an information-rich environment, is a source for both active information seeking and information encountering, which can occur through browsing or as a part of active seeking.

Pregnancy loss–related information seekers were similar to other health information seekers [40,41] in sharing high regard for information from medical professionals but needing additional information as well. Medical professionals are unable to meet certain personal needs that only knowledge gained from lived experiences can provide [80,81]. The main types of health information sought in prior literature in other health contexts [32,33] align with our participants' expressed need for evidence- and experience-based information. Participants did not expect peers to act as pseudo–medical professionals providing medical expertise or interpreting medical advice but instead to offer suggestions on topics such as how to improve communication with clinicians or insight into experiences with a recommended treatment or procedure [80,81]. We note that in contrast to the documented lack of trust between individuals who are socially marginalized or stigmatized and health care providers [12,14,15], participants expressed trust in medical professionals; their concerns centered on not receiving enough information on this particular topic and possibly not knowing how to broach it with medical professionals. Nevertheless, this expression of trust in medical professionals could be because of our sample limitations as most participants were White and of higher socioeconomic status.

Even so, we extend the scholarship on pregnancy-related information behavior. Participants expressed a desire for information on pregnancy loss *early on* in the pregnancy to prepare for the possibility of complications arising, which resonates with earlier research demonstrating that individuals with pregnancies begin their search for information at the outset of the process [53]; however, research on information needs after pregnancy loss suggests that most seek this information after the loss, not before it [28]. Similar to prior research on the information-seeking (not accounting for loss) [53,82] of individuals with pregnancies, participants in our study showed a propensity for changing behavior in response to the information they discovered on internet-based spaces. Participants also demonstrated a tendency to use this information to advocate for themselves in clinical encounters, which has not been heavily reported on in expectant parents whose pregnancies do not lead to a loss [54].

Some barriers for participants were similar to those faced by other pregnancy-related information seekers: not having instruction on or assistance in conveying information to a partner and lacking targeted information for individual experiences

[76]. Gold et al [61] were one of the first to examine support seeking for pregnancy loss on the internet; our findings reiterate the important role that internet-based support and resources can play. Participants in the study by Gold et al [61] also expressed a need for additional evidence-based information, feeling validated in internet-based spaces, appreciating the ready information access that internet-based spaces can provide, and the value of seeing others move on after pregnancy loss or related medical issues. However, Gold et al [61] focused on the benefits of these internet-based spaces and did not address the barriers to acquiring pregnancy loss–related information or their associated outcomes. Our findings on others' experiences and medical providers as sources of information also resonate with the study by Kresnye et al [28]. Although Kresnye et al [28] explored challenges in accessing information, such as locating resources and self-blame, we identified other challenges with respect to accessing, understanding, and applying information related to the technical features required for effective search and assessing information credibility.

We discuss internet-based health literacy concerns and considerations for design in the following section. Pregnancy-related apps generally do not account for pregnancy loss [17], including information behavior–related needs. We suggest that future pregnancy-related and pregnancy loss–related information spaces should address the needs and challenges that we identify herein [28].

Designing for Assessability, Searchability, and Persistence

Overview

An important aspect of the internet-based health literacy process is assessing data quality within an information environment to determine its usefulness [83,84]. Our participants described internet-based health literacy and data quality as concerns in connection to credibility challenges when seeking information.

The themes we identified resonated with the need to design for populations based on their specific information needs and barriers to meeting those needs, as emphasized in current scholarship on health information behavior in internet-based spaces. Individuals who have experienced pregnancy loss have information needs that, although specific to them, also share the concerns over determining credibility with others [21,85]. Our themes also resonated with current HCI literature on the needs of patients managing chronic disease in internet-based health and support groups [80,81], including consideration of the user as a whole, personal and clinical information, clear organization and the ability to search archived information, and safeguards against misinformation. The remainder of this section addresses these concerns.

Designing for Assessability

Our findings speak to the need to extend the *assessable design* framework [86] to the internet-based health context. Designing for assessability was first introduced by Forte et al [86], fusing information literacy concepts in various disciplines toward developing a framework for the *assessable design* of participatory information sources (eg, Wikipedia, forums, and support groups). The authors established two concepts critical

for assessability by reviewing the information credibility literature: (1) information provenance, concerned with determining where information comes from, and (2) information stewardship, which refers to how an information space is maintained [86]. Assessable designs should facilitate the understanding of how the information in internet-based participatory environments is produced, how an environment is sustained, and how to contribute to the space [86]. In fact, prior research investigating the processes individuals use to determine credibility found that users are likely to refer to the source; sources are deemed more credible if the poster is a professional expert or an expert according to community status or past engagement [87-89].

Applying the assessable design framework to our results, internet-based spaces could provide users with affordances that allow them to identify and assess information sources and contributors and their credibility; for example, internet-based groups may consider adding identity tags to content producers that would enable others to assess their contributions. These recommendations resonate with prior suggestions for formal patient–provider internet-based spaces for patients with chronic diseases [80,81]; our findings extend these design needs into informal digital spaces and beyond chronic disease management to individuals managing experiences such as pregnancy loss that can be acute or chronic.

Participants expressed the need for information in two categories: others' stories and experiences (experience-based) and scientifically oriented information (evidence-based). Individuals want to know that their information source is a knowledgeable expert, whether because of training or lived experience. Therefore, designing for the clear identification of the source can assist users in determining the credibility and usefulness of the information, as people are more likely to adjust their attitudes or behaviors in response to a message from someone deemed an expert [89]. These design directions could assist in users' effectiveness and confidence in doing what we found they already try to do: advocate for themselves with medical professionals.

Although identity attributes should be developed together with relevant stakeholders in future work, examples could include determining whether a person has relevant professional expertise. Encouraging or enforcing the citation of sources for the information shared in internet-based spaces as part of the community guidelines is another consideration; individuals could be prompted to add references to the content they share so that others can assess those sources and the information's credibility.

Similar to Forte et al [86], internet-based spaces could also provide users with aggregate information (eg, through simple visualizations) that show what sources contributors draw from. For example, one may go to a Facebook pregnancy loss group and see that 70% of posts link to outside sources, and out of those, 30% are academic peer-reviewed articles about pregnancy loss. This would likely provide the user with an assessment of the group as a whole. We advocate for including assessability in design considerations.

Designing for Searchability and Persistence

Our analysis led to identifying searchability and persistence as desired affordances for participants. For example, some participants shared successful experiences with a Facebook group's search feature and less successful experiences with some pregnancy-related apps and forums. The point here is not to compare platforms (which requires other methods and is an area for future work) but to learn from participants' experiences. Our findings show that people experiencing pregnancy loss need both evidence-based and experience-based information. There is a wide range of adjacent medical conditions (eg, polycystic ovary syndrome) that shape people's loss experiences; finding information about those conditions was also important to participants. Taken together, future designs could experiment with features that allow people to tag the content they contribute as *evidence-based* or *experience-based* or with adjacent health condition labels to streamline searching for others. Although this needs further exploration, this design application would likely be well-suited for an integrated patient–provider application, as described in the study by Huh et al [81], where peers can share and discuss content with one another and consult a medical professional as needed. A possible added benefit of such a design could be the ability for the expert medical professional to share insights without compromising peer-to-peer sharing, which was an issue observed by Huh et al [81].

Such an approach could also be used to provide aggregate information about what experiences and information types are represented within an internet-based space (eg, 79% experience-based). This design approach could be combined with the approach suggested by Huh et al [81], which calls for designing systems that can suggest prior threads relevant to a user's post, or with the suggestion of Hartzler et al [80] for profile features that detail topics a user typically discusses or posts information about.

In summary, our findings show that when there is a dire need for information, access to relevant experiences and information and the ability to assess the credibility of such information are key. The first can be supported by designing for easy retrieval and organized storage of older content through searchability and persistence. The second could be supported by an assessable design. That said, privacy considerations are crucial; for example, if contributors want their data to expire after a certain point, they should be able to easily achieve that goal. Privacy concerns are especially important when designing for information retrieval in the context of well-being in internet-based communities, which are not regulated at the same level as formal, traditional health information (eg, electronic health records). Altogether, we argue that designing for assessability, persistence, and searchability is needed to achieve dimensions of health literacy (ie, access, understand, appraise, and apply) [84]. Without these aspects of design, individuals are less equipped to find information, make informed decisions about how the information could assist in maintaining their health, and effectively communicate this information to others. Researchers and technologists could consider these recommendations when designing pregnancy-related internet-based spaces, which tend to grossly neglect pregnancy loss [17], as well as spaces dedicated to pregnancy loss. Accounting for pregnancy loss in designing internet-based information spaces will help counter the stigma many endure and make space for diverse pregnancy narratives.

Limitations

The larger context within which survivors experience pregnancy loss shapes their experiences. Therefore, we focused primarily on the United States to examine pregnancy loss experiences in this study. We encourage researchers to explore similar topics in other countries. We hoped to include a wide range of experiences (eg, technology use and pregnancy history) and demographics such as age, race, and income level to the extent possible. We were not entirely successful in achieving diversity on all these levels. Although our sample represents a range of technology experience and pregnancy history, it was primarily White, cisgender, heterosexual, married, educated, and urban and had an income >US $75,000. Our sample's demographics are a limitation; however, the findings advance our knowledge about information behavior after pregnancy loss. In the future, we hope to build on this work by reaching more diverse populations, for example, by partnering with community organizations serving marginalized groups and including partners or caretakers. Although an accepted recruitment practice for engaging hard to reach populations in research [68-70], the social media recruitment strategy has its limitations (eg, not everyone who would have been eligible saw our study call).

Our findings are not generalizable; however, they still hold value and contribute to our knowledge about social technologies' roles in information behavior after pregnancy loss. In addition, rather than achieving validity through quantity, in-depth long interview studies with small sample sizes support interpretive claims achieved through the careful selection of participants who share experiences related to research questions [90]. As such, our findings are not intended to be generalizable; rather, they are presented as generative points to provide a conceptual vocabulary for describing information behavior processes. Future work may use representative samples and surveys to assess the prevalence of the identified themes in this paper.

Conflicts of Interest

None declared.

XSL•FO

RenderX

References

1. Wilson T. Information behaviour: an interdisciplinary perspective. Inf Process Manag 1997 Jul;33(4):551-572. [doi: 10.1016/S0306-4573(97)00028-9]

2. Hakimi D, Bryant-Davis T, Ullman SE, Gobin RL. Relationship between negative social reactions to sexual assault disclosure and mental health outcomes of Black and White female survivors. Psychol Trauma 2018 May;10(3):270-275 [FREE Full text] [doi: 10.1037/tra0000245] [Medline: 27936851]

3. Costello KL. Social relevance assessments for virtual worlds. J Document 2017 Oct 04;73(6):1209-1227. [doi: 10.1108/jd-07-2016-0096]

4. The internet and health. Pew Research Center. URL: https://www.pewresearch.org/internet/2013/02/12/the-internet-and-health/ [accessed 2022-02-06]

5. Chen AT. The relationship between health management and information behavior over time: a study of the illness journeys of people living with fibromyalgia. J Med Internet Res 2016 Oct 25;18(10):e269 [FREE Full text] [doi: 10.2196/jmir.5309] [Medline: 27780794]

6. Rubenstein EL. Rituals of introduction and revolving roles: socialization in an online breast cancer community. Library Inf Sci Res 2015 Oct;37(4):353-362. [doi: 10.1016/j.lisr.2015.11.005]

7. Rubenstein EL. "They are always there for me": the convergence of social support and information in an online breast cancer community. J Assn Inf Sci Tec 2014 Nov 06;66(7):1418-1430. [doi: 10.1002/asi.23263]

8. Hoeber O, Patel D, Storie D. A study of academic search scenarios and information seeking behaviour. In: Proceedings of the 2019 Conference on Human Information Interaction and Retrieval. 2019 Presented at: CHIIR '19: Conference on Human Information Interaction and Retrieval; Mar 10 - 14, 2019; Glasgow Scotland UK. [doi: 10.1145/3295750.3298943]

9. Rha E. Exploring the effects of social contexts on task-based information seeking behavior. In: Proceedings of the 2018 Conference on Human Information Interaction & Retrieval. 2018 Presented at: CHIIR '18: Conference on Human Information Interaction and Retrieval; Mar 11 - 15, 2018; New Brunswick NJ USA. [doi: 10.1145/3176349.3176356]

10. Rhee HL. Unique qualities of historians' information-seeking behavior in historical research. Proc Am Soc Inf Sci Technol 2011 Feb 03;47(1):1-2. [doi: 10.1002/meet.14504701362]

11. Chen AT. Information use and illness representations: understanding their connection in illness coping. J Assn Inf Sci Tec 2014 Apr 16;66(2):340-353. [doi: 10.1002/asi.23173]

12. Bronstein J, Knoll M. Blogging motivations of women suffering from infertility. Inf Res 2015 Jun;20(2):1-14 [FREE Full text]

13. Knoll M, Bronstein J. An examination of the information disclosure behavior of infertility bloggers: patterns of self-disclosure and anonymity. Aslib J Inf Manag 2014 Mar 10;66(2):201. [doi: 10.1108/ajim-06-2013-0055]

14. Naveh S, Bronstein J. Sense making in complex health situations. Aslib J Inf Manag 2019 Nov 18;71(6):789-805. [doi: 10.1108/ajim-02-2019-0049]

15. Communicating Pregnancy Loss: Narrative as a Method for Change. Bern, Switzerland: Peter Lang; 2014.

16. Layne LL. Pregnancy loss, stigma, irony, and masculinities: reflections on and future directions for research on religion in the global practice of IVF. Cult Med Psychiatry 2006 Dec;30(4):537-545. [doi: 10.1007/s11013-006-9034-y] [Medline: 17111208]

17. Andalibi N. Symbolic annihilation through design: pregnancy loss in pregnancy-related mobile apps. New Media Soc 2021 Feb 05;23(3):613-631. [doi: 10.1177/1461444820984473]

18. Lupattelli A, Picinardi M, Einarson A, Nordeng H. Health literacy and its association with perception of teratogenic risks and health behavior during pregnancy. Patient Educ Couns 2014 Aug;96(2):171-178. [doi: 10.1016/j.pec.2014.04.014] [Medline: 24862909]

19. Vervier L, Calero VA, Ziefle M. "should I trust or should I go?" or what makes health-related websites appear trustworthy? - an empirical approach of perceived credibility of digital health information and the impact of user diversity. In: Proceedings of the 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health - ICT4AWE. 2018 Presented at: Proceedings of the 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health - ICT4AWE; Mar 22–23, 2018; Funchal, Madeira, Portugal. [doi: 10.5220/0006734401690177]

20. Rooks RN, Wiltshire JC, Elder K, BeLue R, Gary LC. Health information seeking and use outside of the medical encounter: is it associated with race and ethnicity? Soc Sci Med 2012 Jan;74(2):176-184. [doi: 10.1016/j.socscimed.2011.09.040] [Medline: 22154611]

21. Greyson D. Information triangulation: a complex and agentic everyday information practice. J Assoc Inf Sci Technol 2018 Feb 27;69(7):869-878. [doi: 10.1002/asi.24012]

22. Wiltshire J, Cronin K, Sarto G, Brown R. Self-advocacy during the medical encounter: use of health information and racial/ethnic differences. Med Care 2006 Feb;44(2):100-109. [doi: 10.1097/01.mlr.0000196975.52557.b7] [Medline: 16434908]

23. Valdiserri RO, Holtgrave DR, Poteat TC, Beyrer C. Unraveling health disparities among sexual and gender minorities: a commentary on the persistent impact of stigma. J Homosex 2019;66(5):571-589. [doi: 10.1080/00918369.2017.1422944] [Medline: 29297774]

XSL·FO
RenderX

24. Singh A, Gibbs J, Blandford A. Emotion and experience in negotiating HIV-related digital resources: "It's not just a runny nose!". In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow Scotland UK. [doi: 10.1145/3290605.3300829]

25. Andalibi N. What happens after disclosing stigmatized experiences on identified social media: individual, dyadic, and social/network outcomes. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow Scotland UK. [doi: 10.1145/3290605.3300367]

26. Andalibi N. Disclosure, privacy, and stigma on social media. ACM Trans Comput Human Interact 2020 Jun 16;27(3):1-43. [doi: 10.1145/3386600]

27. Andalibi N, Forte A. Announcing pregnancy loss on Facebook: a decision-making framework for stigmatized disclosures on identified social network sites. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18: CHI Conference on Human Factors in Computing Systems; Apr 21 - 26, 2018; Montreal QC Canada. [doi: 10.1145/3173574.3173732]

28. Kresnye K, Alqassim M, Hollins B, Guerra-Reyes L, Wolters M, Siek K. What to expect when you are no longer expecting: information needs of women who experienced a miscarriage. In: Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2020 Presented at: PervasiveHealth '20: Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 18 - 20, 2020; Atlanta GA USA. [doi: 10.1145/3421937.3421995]

29. Hardy S, Kukla R. Making sense of miscarriage online. J Soc Philos 2015 Mar 17;46(1):106-125. [doi: 10.1111/josp.12089]

30. Andalibi N, Garcia P. Sensemaking and coping after pregnancy loss: the seeking and disruption of emotional validation online. Proc ACM Human Comput Interact 2021 Apr;5(CSCW1):1-32. [doi: 10.1145/3449201]

31. Kari J. Diversity in the conceptions of information use. Inf Res 2010 Sep;15(3) [FREE Full text]

32. Fisher S. Meeting between experts: an approach to sharing ideas in medical consultations (book). Soc Health Illness 1987 Jun;9(2):211-213 [FREE Full text] [doi: 10.1111/1467-9566.ep11346994]

33. Song H, Omori K, Kim J, Tenzek KE, Morey Hawkins J, Lin W, et al. Trusting social media as a source of health information: online surveys comparing the United States, Korea, and Hong Kong. J Med Internet Res 2016 Mar 14;18(3):e25 [FREE Full text] [doi: 10.2196/jmir.4193] [Medline: 26976273]

34. Erdelez S. Information encountering: a conceptual framework for accidental information discovery. Inf Encountering 1997 [FREE Full text] [doi: 10.1108/al.1999.8.5.184.9]

35. Toms E. Understanding and facilitating the browsing of electronic text. Int J Human Comput Stud 2000 Mar;52(3):423-452. [doi: 10.1006/ijhc.1999.0345]

36. Williamson K. Discovered by chance: the role of incidental information acquisition in an ecological model of information use. Library Inf Sci Res 1998 Jan;20(1):23-40. [doi: 10.1016/s0740-8188(98)90004-4]

37. Zhang M, Jansen B. Influences of mood on information seeking behavior. In: Proceedings of the CHI '09 Extended Abstracts on Human Factors in Computing Systems. 2009 Presented at: CHI '09: CHI Conference on Human Factors in Computing Systems; Apr 4 - 9, 2009; Boston MA USA. [doi: 10.1145/1520340.1520492]

38. Gaur H, Sharma R. Information seeking behaviour of users in university libraries in digital environment. In: Proceedings of the International Conference and Workshop on Emerging Trends in Technology. 2010 Presented at: ICWET '10: International Conference and Workshop on Emerging Trends in Technology; Feb 26 - 27, 2010; Mumbai Maharashtra India. [doi: 10.1145/1741906.1742038]

39. Lambert SD, Loiselle CG. Health information seeking behavior. Qual Health Res 2007 Oct;17(8):1006-1019. [doi: 10.1177/1049732307305199] [Medline: 17928475]

40. Fox S. The social life of health information. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/ [accessed 2022-02-08]

41. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004 Nov;59(9):1795-1806. [doi: 10.1016/j.socscimed.2004.02.020] [Medline: 15312915]

42. Lee C, Ramírez AS, Lewis N, Gray SW, Hornik RC. Looking beyond the internet: examining socioeconomic inequalities in cancer information seeking among cancer patients. Health Commun 2012;27(8):806-817 [FREE Full text] [doi: 10.1080/10410236.2011.647621] [Medline: 22356137]

43. Andalibi N, Haimson OL, Choudhury MD, Forte A. Social support, reciprocity, and anonymity in responses to sexual abuse disclosures on social media. ACM Trans Comput Human Interact 2018 Oct 31;25(5):1-35. [doi: 10.1145/3234942]

44. Andalibi N, Haimson O, De Choudhury M, Forte A. Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI'16: CHI Conference on Human Factors in Computing Systems; May 7 - 12, 2016; San Jose California USA. [doi: 10.1145/2858036.2858096]

45. Zhou L, Zhang D, Yang C, Wang Y. Harnessing social media for health information management. Electron Commer Res Appl 2018;27:139-151 [FREE Full text] [doi: 10.1016/j.elerap.2017.12.003] [Medline: 30147636]

46. Saparova D, Williams JA, Inabnit CK, Fiesta M. Information behavior shift: how and why medical students use Facebook. Proc Am Soc Info Sci Tech 2014 May 08;50(1):1-4. [doi: [10.1002/meet.14505001171](#)]

47. Paul S, Reddy M. Understanding together: sensemaking in collaborative information seeking. In: Proceedings of the 2010 ACM conference on Computer supported cooperative work. 2010 Presented at: CSCW '10: Computer Supported Cooperative Work; Feb 6 - 10, 2010; Savannah Georgia USA. [doi: [10.1145/1718918.1718976](#)]

48. Minett JW, Zheng H, Fong MC, Zhou L, Peng G, Wang WS. A chinese text input brain–computer interface based on the P300 speller. Int J Human Comput Interact 2012 Jul;28(7):472-483. [doi: [10.1080/10447318.2011.622970](#)]

49. Zhang Y. Proceedings of the 1st ACM International Health Informatics Symposium. 2010 Presented at: IHI '10: Proceedings of the 1st ACM International Health Informatics Symposium; Nov 11 - 12, 2010; Arlington Virginia USA. [doi: [10.1145/1882992.1883023](#)]

50. Huang M, Hansen D, Xie B. Older adults' online health information seeking behavior. In: Proceedings of the 2012 iConference. 2012 Presented at: iConference 2012; Feb 7-10, 2012; Toronto, Ontario, Canada. [doi: [10.1145/2132176.2132220](#)]

51. Guerra-Reyes L, Christie VM, Prabhakar A, Harris AL, Siek KA. Postpartum health information seeking using mobile phones: experiences of low-income mothers. Matern Child Health J 2016 Nov;20(Suppl 1):13-21 [[FREE Full text](#)] [doi: [10.1007/s10995-016-2185-8](#)] [Medline: [27639571](#)]

52. Guerra-Reyes L, Christie VM, Prabhakar A, Siek KA. Mind the gap: assessing the disconnect between postpartum health information desired and health information received. Womens Health Issues 2017;27(2):167-173. [doi: [10.1016/j.whi.2016.11.004](#)] [Medline: [28063847](#)]

53. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016 Mar 28;16:65 [[FREE Full text](#)] [doi: [10.1186/s12884-016-0856-5](#)] [Medline: [27021727](#)]

54. Declercq ER, Sakala C, Corry MP, Applebaum S. Listening to mothers II: report of the second national U.S. survey of women's childbearing experiences: conducted January-February 2006 for childbirth connection by Harris interactive(R) in partnership with Lamaze international. J Perinat Educ 2007;16(4):15-17 [[FREE Full text](#)] [doi: [10.1624/105812407X244778](#)] [Medline: [18769522](#)]

55. Lagan BM, Sinclair M, Kernohan WG. Internet use in pregnancy informs women's decision making: a web-based survey. Birth 2010 Jun;37(2):106-115. [doi: [10.1111/j.1523-536X.2010.00390.x](#)] [Medline: [20557533](#)]

56. Zhu C, Zeng R, Zhang W, Evans R, He R. Pregnancy-related information seeking and sharing in the social media era among expectant mothers: qualitative study. J Med Internet Res 2019 Dec 04;21(12):e13694 [[FREE Full text](#)] [doi: [10.2196/13694](#)] [Medline: [31799939](#)]

57. Gui X, Chen Y, Kou Y, Pine K, Chen Y. Investigating support seeking from peers for pregnancy in online health communities. Proc ACM Human Comput Interact 2017 Dec 06;1(CSCW):1-19. [doi: [10.1145/3134685](#)]

58. Burleson G, Naseem M, Toyama K. An exploration of African-American pregnant women's information-seeking behavior in Detroit. In: Proceedings of the 2020 International Conference on Information and Communication Technologies and Development. 2020 Presented at: ICTD2020: Proceedings of the 2020 International Conference on Information and Communication Technologies and Development; Jun 17 - 20, 2020; Guayaquil Ecuador. [doi: [10.1145/3392561.3394647](#)]

59. Sajjad U, Shahid S. Baby+: a mobile application to support pregnant women in Pakistan. In: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct. 2016 Presented at: MobileHCI '16: 18th International Conference on Human-Computer Interaction with Mobile Devices and Services; Sep 6 - 9, 2016; Florence Italy. [doi: [10.1145/2957265.2961856](#)]

60. Smith W, Wadley G, Daly O, Webb M, Hughson J, Hajek J, et al. Designing an app for pregnancy care for a culturally and linguistically diverse community. In: Proceedings of the 29th Australian Conference on Computer-Human Interaction. 2017 Presented at: OzCHI '17: 29th Australian Conference on Human-Computer Interaction; Nov 28 - Dec 1, 2017; Brisbane Queensland Australia. [doi: [10.1145/3152771.3152807](#)]

61. Gold KJ, Boggs ME, Mugisha E, Palladino CL. Internet message boards for pregnancy loss: who's on-line and why? Womens Health Issues 2012 Jan;22(1):e67-e72 [[FREE Full text](#)] [doi: [10.1016/j.whi.2011.07.006](#)] [Medline: [21907592](#)]

62. Savitz DA, Hertz-Picciotto I, Poole C, Olshan AF. Epidemiologic measures of the course and outcome of pregnancy. Epidemiol Rev 2002;24(2):91-101. [doi: [10.1093/epirev/mxf006](#)] [Medline: [12762085](#)]

63. Fairchild J. What might have been: the communication of social support and women's post-miscarriage narrative reconstruction. ProQuest. URL: http://www.proquest.com/docview/858610979/abstract/99F77901C7344D1APQ/1 [accessed 2022-02-08]

64. Andalibi N, Morris ME, Forte A. Testing waters, sending clues: indirect disclosures of socially stigmatized experiences on social media. Proc ACM Human Comput Interact 2018 Nov;2(CSCW):1-23. [doi: [10.1145/3274288](#)]

65. Andalibi N, Forte A. Responding to sensitive disclosures on social media. ACM Trans Comput Human Interact 2018 Dec 20;25(6):1-29. [doi: [10.1145/3241044](#)]

66. Moulder C. Towards a preliminary framework for understanding pregnancy loss. J Reproductive Infant Psychol 2007 Dec 11;12(1):65-67. [doi: [10.1080/02646839408408869](#)]

67. Andrade C. The inconvenient truth about convenience and purposive samples. Indian J Psychol Med 2021 Jan;43(1):86-88 [[FREE Full text](#)] [doi: [10.1177/0253717620977000](#)] [Medline: [34349313](#)]

68. Martinez O, Wu E, Shultz AZ, Capote J, López Rios J, Sandfort T, et al. Still a hard-to-reach population? Using social media to recruit Latino gay couples for an HIV intervention adaptation study. J Med Internet Res 2014 Apr 24;16(4):e113 [FREE Full text] [doi: 10.2196/jmir.3311] [Medline: 24763130]

69. Shere M, Zhao XY, Koren G. The role of social media in recruiting for clinical trials in pregnancy. PLoS One 2014;9(3):e92744 [FREE Full text] [doi: 10.1371/journal.pone.0092744] [Medline: 24671210]

70. Gelinas L, Pierce R, Winkler S, Cohen IG, Lynch HF, Bierer BE. Using social media as a research recruitment tool: ethical issues and recommendations. Am J Bioeth 2017 Mar;17(3):3-14 [FREE Full text] [doi: 10.1080/15265161.2016.1276644] [Medline: 28207365]

71. Snyder J. Visualizing personal rhythms: a critical visual analysis of mental health in flux. In: Proceedings of the 2020 ACM Designing Interactive Systems Conference. 2020 Presented at: DIS '20: Designing Interactive Systems Conference 2020; Jul 6 - 10, 2020; Eindhoven Netherlands. [doi: 10.1145/3357236.3395463]

72. Zolyomi A, Snyder J. Social-emotional-sensory design map for affective computing informed by neurodivergent experiences. Proc ACM Human Comput Interact 2021 Apr 13;5(CSCW1):1-37. [doi: 10.1145/3449151]

73. Dettmer S, Simpson RL, Myles BS, Ganz JB. The use of visual supports to facilitate transitions of students with autism. Focus Autism Other Dev Disabl 2016 Sep 14;15(3):163-169. [doi: 10.1177/108835760001500307]

74. Prabhakar A, Guerra-Reyes L, Effron A, Kleinschmidt V, Driscoll M, Peters C, et al. "Let me know if you need anything": support realities of new mothers. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2017 Presented at: PervasiveHealth '17: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 23 - 26, 2017; Barcelona Spain. [doi: 10.1145/3154862.3154863]

75. Strauss A, Corbin J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory 3rd edition. Thousand Oaks, California: Sage Publications; 2008.

76. Peyton T, Poole E, Reddy M, Kraschnewski J, Chuang C. Information, sharing and support in pregnancy: addressing needs for mHealth design. In: Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing. 2014 Presented at: CSCW'14: Computer Supported Cooperative Work; Feb 15 - 19, 2014; Baltimore Maryland USA. [doi: 10.1145/2556420.2556489]

77. DeVito M, Birnholtz J, Hancock J. Platforms, people, and perception: using affordances to understand self-presentation on social media. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: CSCW '17: Computer Supported Cooperative Work and Social Computing; Feb 25- Mar 1, 2017; Portland Oregon USA. [doi: 10.1145/2998181.2998192]

78. Preece J, Nonnecke B, Andrews D. The top five reasons for lurking: improving community experiences for everyone. Comput Human Behav 2004 Mar;20(2):201-223. [doi: 10.1016/j.chb.2003.10.015]

79. Zimmerman MS. Information poverty and reproductive healthcare: assessing the reasons for inequity between income groups. Soc Work Public Health 2017 Apr 03;32(3):210-221. [doi: 10.1080/19371918.2016.1268990] [Medline: 28129076]

80. Hartzler A, Pratt W. Managing the personal side of health: how patient expertise differs from the expertise of clinicians. J Med Internet Res 2011 Aug 16;13(3):e62 [FREE Full text] [doi: 10.2196/jmir.1728] [Medline: 21846635]

81. Huh J, Patel R, Pratt W. Tackling dilemmas in supporting 'the whole person' in online patient communities. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI '12: CHI Conference on Human Factors in Computing Systems; May 5 - 10, 2012; Austin Texas USA. [doi: 10.1145/2207676.2208535]

82. Wallwiener S, Müller M, Doster A, Laserer W, Reck C, Pauluschke-Fröhlich J, et al. Pregnancy eHealth and mHealth: user proportions and characteristics of pregnant women using web-based information sources-a cross-sectional study. Arch Gynecol Obstet 2016 Nov;294(5):937-944. [doi: 10.1007/s00404-016-4093-y] [Medline: 27084763]

83. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]

84. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]

85. Rose ID, Friedman DB, Spencer SM, Annang L, Lindley LL. Health information–seeking practices of African American young men who have sex with men. Youth Soc 2013 Jun 20;48(3):344-365. [doi: 10.1177/0044118x13491769]

86. Forte A, Andalibi N, Park T, Willever-Farr H. Designing information savvy societies: an introduction to assessability. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI '14: CHI Conference on Human Factors in Computing Systems; Apr 26 - May 1, 2014; Toronto Ontario Canada. [doi: 10.1145/2556288.2557072]

87. Petcharanan Y, Chartprasert D. Credibility judgment of health information on Facebook: a case of Thai Facebook users. BU Acad Rev 2020;19(1):16 [FREE Full text]

88. Thon FM, Jucks R. Believing in expertise: how authors' credentials and language use influence the credibility of online health information. Health Commun 2017 Jul;32(7):828-836. [doi: 10.1080/10410236.2016.1172296] [Medline: 27466693]

89. Yang Q, Beatty M. A meta-analytic review of health information credibility: belief in physicians or belief in peers? Health Inf Manag 2016 Aug;45(2):80-89. [doi: 10.1177/1833358316639432] [Medline: 27105476]

90. Crouch M, McKenzie H. The logic of small samples in interview-based qualitative research. Soc Sci Inf 2016 Jun 29;45(4):483-499. [doi: 10.1177/0539018406069584]

Abbreviations

HCI: human-computer interaction
LGBTQ: lesbian, gay, bisexual, transgender, and queer or questioning

XSL•FO
RenderX

<u>Original Paper</u>

Readability of Korean-Language COVID-19 Information from the South Korean National COVID-19 Portal Intended for the General Public: Cross-sectional Infodemiology Study

Hana Moon[1], MD, PhD; Geon Ho Lee[1], MSc, MD; Yoon Jeong Cho[1], MD, PhD

Department of Family Medicine, Daegu Catholic University School of Medicine, Daegu, Republic of Korea

Corresponding Author:
Geon Ho Lee, MSc, MD
Department of Family Medicine
Daegu Catholic University School of Medicine
33, Duryugongwon-ro 17-gil, Nam-gu
Daegu, 42472
Republic of Korea
Phone: 82 53 650 4696
Email: totoslee@cu.ac.kr

Abstract

Background: The coronavirus pandemic has increased reliance on the internet as a tool for disseminating information; however, information is useful only when it can be understood. Prior research has shown that web-based health information is not always easy to understand. It is not yet known whether the Korean-language COVID-19 information from the internet is easy for the general public to understand.

Objective: We aimed to evaluate the readability of Korean-language COVID-19 information intended for the general public from the national COVID-19 portal of South Korea.

Methods: A total of 122 publicly available COVID-19 information documents written in Korean were obtained from the South Korean national COVID-19 portal. We determined the level of readability (at or below ninth grade, 10th to 12th grade, college, or professional) of each document using a readability tool for Korean-language text. We measured the reading time, character count, word count, sentence count, and paragraph count for each document. We also evaluated the characteristics of difficult-to-read documents to modify the readability from difficult to easy.

Results: The median readability level was at a professional level; 90.2% (110/122) of the information was difficult to read. In all 4 topics, few documents were easy to read (overview: 5/12, 41.7%; prevention: 6/97, 6.2%; test: 0/5, 0%; treatment: 1/8, 12.5%; P=.006), with a median 11th-grade readability level for overview, a median professional readability level for prevention, and median college readability levels for test and treatment. Difficult-to-read information had the following characteristics in common: literacy style, medical jargon, and unnecessary detail.

Conclusions: In all 4 topics, most of the Korean-language COVID-19 web-based information intended for the general public provided by the national COVID-19 portal of South Korea was difficult to read; the median readability levels exceeded the recommended ninth-grade level. Readability should be a key consideration in developing public health documents, which play an important role in disease prevention and health promotion.

Introduction

Digital health literacy is the ability to seek, find, understand, and appraise health information from electronic sources, and the subsequent ability to apply the knowledge to address a health problem [1]. Many people have difficulty understanding written information worldwide, including approximately 9.6 million Koreans [2] and approximately 75 million Americans [3]. Therefore, providing health information without considering

the population's literacy level does not guarantee people's ability to understand and apply the information [4].

Society should ensure that vulnerable individuals are not left behind. Health literacy is a social determinant of health—people with lower literacy skills are less likely to access information or health care services at the same level as those with higher literacy skills [5]. This may contribute to poor health outcomes, such as lower adherence to infection control and prevention measures, ineffective use of health care, and high mortality rates [5-7]. In this context, public health researchers and policy makers have recently extended the concept of health literacy from personal reading skills to organizational health literacy [5,8,9]. The US Department of Health and Human Services' health strategy [10] emphasized that it is the responsibility of health care organizations to design and deliver health care services and information relating to health in an accessible and understandable format. The South Korean government also includes developing easy-to-read health information in its action plan to reduce health inequity [11]. These public health efforts to achieve health equity will reduce the risk of increasing health disparities within and between countries [10-13].

Readability refers to how easy a text is to read and understand, and it is commonly measured by school grade level (kindergarten to postgraduate school) [14]. Generally, a text is considered easy to read when written below the average reading level of an adult [14]. Health care authorities have encouraged enhancing the readability of health care information, recommending that health information intended for the public be written below a sixth-grade reading level [15-17]. The South Korean government did not establish a standard for the readability of health information despite doing so for other information documents, with the Easy-to-Understand Legislation Project in 2006, which recommended that documents for the general public be written at or below a ninth-grade reading level [18]. This standard was based on the average reading level of Korean adults and the 9 years of free compulsory education that South Koreans receive [19]. The ninth-grade level was also used as a standard for sufficient literacy skills required for daily life in the 2017 Second Korean Adult Literacy Survey conducted by the Ministry of Education and the National Institute for Continuing Education [2]; thus, similarly, this study considers public health information adequately readable when written at or below that of a ninth-grade reading level.

The COVID-19 pandemic has changed the way we communicate, with more people relying on the internet as their primary source of information [20]. COVID-19–related searches have surged and been predominant in 2020 [21,22]. Consequently, web-based communication regarding the risks of the virus has become increasingly important [23,24]. According to a UN report [25], governments have used websites to provide accurate information for the public since the early days of the COVID-19 pandemic. Although health care's digital transformation has increased the accessibility of information [26], many people may not understand the COVID-19 information shared by health care authorities. Prior research [15,27-34] has shown that most available health information is difficult for the general public to understand. For example, a systematic review [15] of 157 cross-sectional studies concluded

that the US and Canada's web-based health information is written above the average reading level of the population that it aims to inform. Patient information leaflets written in Korean were also found to be written above the average reading level of South Korean adults [27-30], and recently, studies [31-34] have reported that web-based COVID-19 information (written in English) is considered too difficult to understand by the public. However, no studies have been conducted on Korean-language web-based COVID-19 information. We aimed to address this literature gap by evaluating the readability of Korean-language COVID-19 resources. We investigated three research questions: (1) Is Korean-language COVID-19 information provided for the general public on the national COVID-19 portal of South Korea written at or below the recommended ninth-grade level? (2) Does readability differ across topics? (3) What are the characteristics of difficult-to-read information, and how can we improve readability?

Methods

Search Strategy

Information posted between February 3, 2020 and February 10, 2021 was downloaded from the national COVID-19 portal of South Korea [35] on February 10, 2021. Any subsidiary webpages or subdirectories that had information accessible by the public were also assessed using software (Sitechecker, version February 2021; Boosta Inc). All documents were initially screened by title and the main text was reviewed by 2 authors (HM and GHL) independently, using inclusion and exclusion criteria. Any discrepancies were resolved through discussion with the third author (YJC). Information that contained (1) COVID-19 information intended for the general public, (2) was written in Korean-language, and (3) provided by the South Korean government was included. Information that was (1) in other languages, (2) noneducational (such as press releases or daily case updates), (3) not in a written format (ie, videos and images), or (4) intended for public health and health care professionals was excluded. We also excluded duplicate documents.

Ethics

In this study, there were no human participants or assigned interventions; therefore, we did not seek specific ethical approval from an institution review board.

Topic Classification

Included documents were classified by topic—overview, prevention, test, or treatment—by 2 of the authors (HM and GHL) independently, and any disagreement was resolved via discussion with the third author (YJC). The *overview* category included documents about COVID-19 risk factors, transmission, and the natural course of the disease. The *prevention* category included documents discussing cleaning, disinfection, physical distancing, personal protective equipment, and vaccination. The *test* category included documents discussing indications, screening and confirmation tests, or the interpretation of test results. The *treatment* category included documents about self-care and patient care.

XSL•FO
RenderX

Text Preparation

Documents were formatted as raw text files using Notepad (Microsoft Inc). Any text not directly related to public education was deleted, such as date, author information, titles, figures, tables, legends, references, and copyright information [31,33,36].

Readability Assessment

We assessed text readability using a tool for Korean text (Natmal, version 2019; Lexical data-processing research institute) that determines the readability level based on sentence length and word difficulty [37]. The Natmal database contains 500,000 words classified into 9 difficulty levels by grade [38]. The frequency of words used in the text is measured for each grade and then weighted according to the number of words listed in the database [37]. We grouped text into 4 levels: professional, college, 10th to 12th grade, and at or below ninth grade. Difficult-to-read information was defined as information with a readability level exceeding the ninth-grade level (professional, college, 10th to 12th grade), and easy-to-read information was defined as information with a readability level at or below the ninth-grade level [2].

Enhancing Readability

We analyzed the characteristics of difficult-to-read information to determine common characteristics. Out of the documents that were written at a professional level, 3 documents, each representing a characteristic, were selected after discussion among the authors. To modify the readability level, we addressed each problem characteristic. The readability of each revised document was reassessed with the tool.

Statistical Analysis

All analyses were conducted with the R statistical software (version 3.6.3; R Foundation for Statistical Computing). Results were considered significant with $P<.05$. Shapiro-Wilk tests were used to assess the data normality. The categorical variable (readability level) was presented as frequency and percentage, and continuous variables (reading time, character count, word count, sentence count, and paragraph count) were presented as median and interquartile range. Chi-square tests or Fisher exact tests were used to calculate P values for categorical variables. Kruskal-Wallis tests were used to calculate P values for continuous variables.

Results

Readability of the Documents

A total of 122 educational documents were included in this study (Figure 1). The median readability level was *professional*; 9.8% (12/122) of documents were classified as easy to read, and 90.2% (110/122) were classified as difficult to read (Table 1).

Figure 1. Data collection flowchart.
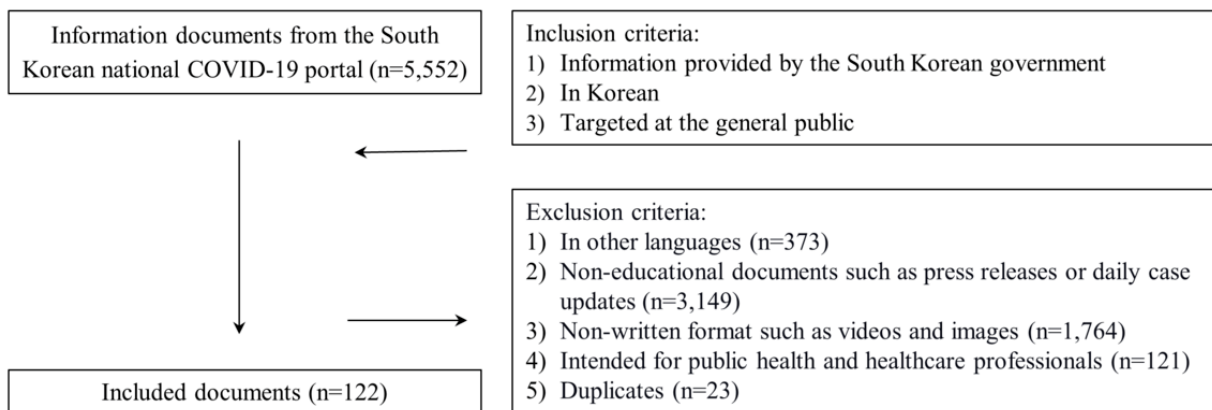
Table 1. Readability characteristics by level.

Characteristic	All (n=122)	Level				P value
		Professional (n=66)	College (n=33)	10th to 12th grade (n=11)	At or below ninth grade (n=12)	
Reading time (seconds), median (IQR)	78.4 (43.0, 161.3)	129.4 (57.2, 191.6)	64.4 (43.9, 107.7)	38.9 (33.4, 93.9)	32.8 (24.9, 48.8)	<.001
Character count, median (IQR)	1350.5 (703.0, 2418.0)	2003.0 (945.0, 2926.0)	1242.0 (739.0, 1837.0)	701.0 (620.5, 1609.0)	632.5 (475.5, 878.0)	<.001
Word count, median (IQR)	264.5 (145.0, 544.0)	436.5 (193.0, 646.0)	217.0 (148.0, 363.0)	131.0 (112.5, 316.5)	110.5 (84.0, 164.5)	<.001
Sentence count, median (IQR)	31.5 (18.0, 60.0)	47.0 (24.0, 68.0)	26.0 (14.0, 36.0)	21.0 (17.5, 36.0)	17.0 (10.0, 21.5)	<.001
Paragraph count, median (IQR)	35.5 (20.0, 84.0)	75.5 (35.0, 121.0)	27.0 (17.0, 41.0)	23.0 (19.5, 33.0)	13.5 (9.5, 17.0)	<.001

Readability Among the Documents of Different Topics

Included documents did not evenly cover the 4 topics (Table 2), with most (97/122, 79.5%) covering the topic *prevention*. For all topics, median readability was classified as difficult (overview: 11th-grade level; prevention: professional level; test: college level; treatment: college level), and there were few (all cases: *P*=.006) easy-to-read documents (overview: 5/12, 41.7%; prevention: 6/97, 6.2%; test: 0/5, 0%; treatment: 1/8, 12.5%).

Table 2. Readability among the documents of different topics.

Characteristic	All (n=122)	Topic				P value
		Overview (n=12)	Prevention (n=97)	Test (n=5)	Treatment (n=8)	
Readability level, median	Professional	11th grade	Professional	College	College	—[a]
Readability level, n (%)						.006
Professional	66 (54.1)	3 (25.0)	59 (60.8)	2 (40.0)	2 (25.0)	
College	33 (27.0)	3 (25.0)	23 (23.7)	3 (60.0)	4 (50.0)	
10th to 12th grade	11 (9.0)	1 (8.3)	9 (9.3)	0 (0.0)	1 (12.5)	
At or below ninth grade	12 (9.8)	5 (41.7)	6 (6.2)	0 (0.0)	1 (12.5)	
Reading time (seconds), median (IQR)	78.4 (43.0, 161.3)	35.4 (27.4, 57.4)	97.6 (49.8, 170.5)	56.9 (32.0, 64.4)	42.1 (29.9, 93.8)	<.001
Character count, median (IQR)	1350.5 (703.0, 2418.0)	663.0 (565.5, 1176.0)	1559.0 (830.0, 2544.0)	1049.0 (609.0, 1331.0)	763.5 (571.5, 1556.5)	.01
Word count, median (IQR)	264.5 (145.0, 544.0)	119.5 (92.5, 193.5)	329.0 (168.0, 575.0)	192.0 (108.0, 217.0)	142.0 (101.0, 316.5)	<.001
Sentence count, median (IQR)	31.5 (18.0, 60.0)	16.0 (10.0, 27.0)	37.0 (20.0, 63.0)	19.0 (11.0, 29.0)	15.0 (11.0, 29.0)	<.001
Paragraph count, median (IQR)	35.5 (20.0, 84.0)	14.0 (10.5, 30.0)	44.0 (23.0, 97.0)	26.0 (16.0, 31.0)	18.5 (11.5, 36.5)	<.001

[a]This comparison was not made.

Enhancing Readability

Difficult-to-read information had 3 characteristics in common. They were written in literacy style, with medical jargon, and with unnecessary detail (Table 3).

Document A was the answer to the question, "Will I catch COVID-19 if I travel on a bus or subway train previously used by a confirmed patient?" It was written in literacy style, making it difficult to read. It was changed from literacy style to colloquial to improve its readability.

Document B was the answer to the question, "What are the symptoms of COVID-19?" The original version described the symptoms at length, making it difficult to identify important information. To make the document easier to understand, we included only typical symptoms for each organ rather than all possible symptoms. The medical jargon used in document B was replaced with common words to improve its readability. Medical terms, such as dyspnea, hemoptysis, emesis, anosmia, and ageusia, were replaced with commonly used words, such as shortness of breath, coughing up blood, vomiting, and loss of smell or taste.

Document C was the answer to the question, "How is the test for COVID-19 done?" The original version used medical jargon and included excessive detail in the explanation. To improve its readability, the content was summarized, and the unnecessarily detailed information included in document C was replaced with information tailored to the general public.

Table 3. Original text [35] and text with revisions to enhance readability.

Document	Original version	Suggested revision
A, literacy style	When riding on public transportation such as buses or subways, if you touch a handle that has been contaminated by the coronavirus, you could become infected with the virus through your eyes, nose, and mouth. However, proper wearing of a mask and hand sanitization can lower the risk of COVID-19 infection.	You can catch the virus through your eyes, nose, or mouth after touching a handle that has the virus on it. However, if you wear a mask and wash your hands, you can lower your chances of contracting COVID-19.
B, medical jargon	Symptoms such as a fever of 37.5 ℃ or higher, cough, dyspnea, chills, muscle pain, headache, sore throat, anosmia, ageusia, pneumonia, fatigue, decreased appetite, phlegm, confusion, dizziness, and a runny or stuffy nose can indicate COVID-19. Hemoptysis, chest pain, conjunctivitis, skin symptoms, or digestive system symptoms such as nausea, emesis, and diarrhea may also indicate COVID-19.	Fever and cough are common. Having COVID-19 can make you feel cold and tired, and you may have difficulty in breathing. You may lose your appetite and have a sore throat, body aches, nausea, or diarrhea. You may not have any sense of smell or taste. Some people may not have any symptoms and feel normal, but they can still spread the virus.
C, unnecessary detail	Samples are collected by physicians, nurses, and medical technicians at designated locations. Upper respiratory tract sampling is mandatory, while lower respiratory tract sampling is optional for patients with sputum. You may experience discomfort or pain while the sample is being collected. The upper respiratory tract sample is a combination of a nasopharyngeal and oropharyngeal swab in one tube. A nasopharyngeal swab involves inserting a cotton swab into the nostril until it reaches the posterior nares. An oropharyngeal swab includes inserting a cotton swab to scrape the inside of the throat. A lower respiratory tract sample is collected by spitting sputum into a container, ensuring that it is not contaminated by other liquids.	You can get tested at screening centers or drive-through facilities. There is a swab and a spit test. In the swab test, a doctor or nurse inserts a cotton swab into your nose or throat. You may feel some discomfort. For the spit test, you are asked to spit thick mucus from your throat into a tube. You will receive your test results through text message.

Discussion

Principal Results and Study Strengths

This study shows that 90.2% of the information available to the public was difficult to read. Of the documents discussing the prevention of COVID-19, only 6.2% (6/97) could be rated as easy to read. This is noteworthy as it shows that very few documents would effectively be able to spread prevention knowledge to the general population. To encourage people to adopt personal protective measures, such as wearing masks and washing hands during the pandemic, the government should prioritize making information on prevention easier to understand. Moreover, easy-to-read documents that were available did not cover all relevant topics equally, such as overview (n=5), prevention (n=6), test (n=0), and treatment (n=1). To make sure that those with lower literacy skills have access to information on public health and safety topics at the same level as people with higher literacy skills.

To the best of our knowledge, this is the first study to evaluate the readability of web-based COVID-19 information written in Korean. Korean is the 20th most spoken language globally, with approximately 82 million speakers [39]. Moreover, South Korea has a large population of older individuals; in 2020, adults aged 65 years and older accounted for 15.7% of the population [40], and it is estimated that South Korea will become a super-aged society in 2025, when the proportion of older adults is expected to reach 20.3% [40]. The proportion is likely to increase to 43.9% by 2060 [40]. Older adults are an important target group when assembling easy-to-understand COVID-19 information because they are at high risk for developing serious complications from COVID-19 [41]. In addition, many older

adults have low literacy skills [42]. For example, 71% of Americans older than 60 years were reported to have difficulties understanding written information [3]. Nearly one-third (31.27%) of Korean adults aged 55 to 65 years could read the words but could not understand sentences or long texts [19]. Older adults with lower literacy skills are more likely to be marginalized by the government's health care or welfare system because they cannot follow instructions for filling out forms.

The documents used in this study, from the national COVID-19 portal run by the government, which is an integrated communication channel that compiles all the COVID-19 information created by various government agencies [35], are likely an accurate reflection of the information currently accessible to the public. The Korean government's message was amplified through social media platforms (such as Twitter and Facebook) and traditional media (TV and newspapers), thus reaching every corner of the country [43].

Advantages and Disadvantages of the National COVID-19 Portal of South Korea for Distributing Information

The national COVID-19 portal of South Korea has several advantages for disseminating information. First, the website is highly accessible; it appears at the top of the first search results page when searching for COVID-19 on 3 major local search platforms: Naver, which has a 68.9% search engine market share in South Korea; Google, which has a 21.4% market share; and Daum, which has a 7.5% market share [44]. This national COVID-19 portal ranked first in web traffic among South Korean government websites [45], and the website ranked 10th in global rankings for health conditions and concerns [46]. Most

health information seekers begin their search activities with search engines [47]; therefore, it is highly likely that Koreans who searched for COVID-19 information on the internet visited the national COVID-19 portal of South Korea. Furthermore, the information on the national COVID-19 portal are convenient for users because they can access any page without logging in, the documents can be downloaded in various formats, and anyone may freely use these documents for public purposes without any copyright restrictions.

Although the national South Korean COVID-19 portal may serve as a key communication platform between health care authorities and the public, it does not have a user-friendly interface. The web pages of the portal are not divided according to the target audience. As a result, medical professionals visiting this website may waste time reading superficial information, and users who are not medical experts may be overwhelmed with unnecessarily detailed explanations. By separating the pages or sections according to audience type (health care workers, the general public, people with low literacy skills), users may then have a more convenient way to access user-friendly information.

Readability of Web-Based COVID-19 Information Generated by Other Public Health Agencies

The websites of the US Centers for Disease Control and Prevention and the National Health Service England have distinct sections for health care workers, the general public, and people with low literacy skills [48-51]. The web page for health care professionals provides COVID-19 training, such as clinical guides for managing cancer patients and patients requiring endoscopy during the COVID-19 pandemic [50]. Their easy-to-read sections provide information on COVID-19 basics, such as advice surrounding staying at home, COVID-19 vaccine during pregnancy, and what to expect after receiving a COVID-19 vaccine [51].

However, recent studies [34,52,53] have shown that some public health agencies have also failed to provide information in an easy-to-read form. Valizadeh-Haghi et al [52] examined the readability of English-language COVID-19 information based on website categories (ie, news, governmental, commercial, organization, educational) and concluded that that the readability levels in all categories exceeded the recommended level and that commercial websites had better readability than governmental websites. Mishra et al [34] reported that the readability of English-language COVID-19 information on 18 government and international public health agency websites did not meet the recommended readability level. Halboub et al [53] investigated 36 Arabic-language websites on COVID-19 and reported that 66.7% of the included websites were easy for the general public to read. Kruse et al [33] conducted a study of COVID-19 information provided by 8 US academic medical centers and reported that 0.7% of information was written at or below the sixth-grade level.

Comparison With Literature on Readability of Outbreak-Related Information

Reading outbreak-related information poses extra challenges to a layperson because of the use of medical jargon [54-56].

Unfamiliar terms, such as *waterborne*, *vertical*, *zoonotic infection*, *herd immunity*, *incubation period*, *cohort isolation*, *outbreak*, *epidemic*, and *pandemic*, are frequently used. Previous studies [57,58] have shown that information on infectious diseases had poor readability; for example, Ebola virus–related information provided by public health agencies was written at readability levels higher than recommended [57], and Basch et al [58] reported that 93% of web-based Zika virus–related information was difficult to read.

Moreover, understanding COVID-19 information is more challenging because of the use of new words and phrases, such as *social distancing*, *un-tact* (noncontact), *new normal*, and *covideo party*. COVID-19 search results are also difficult for the public to understand—only 17.2% of COVID-19–related web pages were at a readable level [32], and Google-searched information regarding COVID-19 (0/150 articles) did not meet recommended readability levels.

Recent readability studies on English texts have been conducted on various types of texts on specific topics such as those used with vaccine clinical trials, privacy policies, and tests. For example, Emanuel and Boyle [59] reported that informed consent texts for COVID-19 vaccine trials were long and difficult to read. Zhang et al [60] reported that explanations of privacy policies of COVID-19 contact-tracing apps were written at readability levels higher than those recommended. Garcia et al [61] investigated the readability of web-based information on COVID-19 testing and reported that only 6 of 50 websites had appropriate readability.

Interestingly, Mishra et al [34] included information written in English posted on the South Korean government website and reported that it was written above the 11th-grade level; however, these results may not truly represent COVID-19 information commonly shared by Koreans, as South Korea is not an English-speaking country. Korean is the only official language of South Korea; paperwork and internet activities in this region are mainly in Korean. Therefore, information in Korean, rather than information in English, should be analyzed to yield results that represent COVID-19 information commonly shared by Korean.

Implications for Practice

To improve the readability of COVID-19 information aimed at the general public, we urge the South Korean government to introduce the following measures. First, the national COVID-19 portal should be organized according to audience type (ie, health care workers, the general public, and those with a low level of education) to optimize the user experience of each type of audience. Second, guidelines on how to draft easy-to-understand health information for Korean speakers should be developed using plain-writing guidelines that include the principles and skills for easy-to-read writing—the target audience should be identified, important points should be prioritized, information should be provided step by step, foreign words should be reduced, short sentences should be used, important topics should be summarized at the end [62],and synonyms should be used to replace medical jargon with everyday words [63,64].

XSL·FO

RenderX

Limitations

Our study has several limitations. First, the results might be biased since only one tool was used to assess readability. Second, our results are representative only of the search time frame. Numerous COVID-19 studies have been conducted, and web-based information related to it is constantly changing [65]. Third, there is no verifiable information regarding who has accessed or read what information in the study because collecting or using any personal information was not allowed. Fourth, we only used the readability tool and did not test any information with actual reader. Fifth, we only analyzed text and did not consider other factors that could affect readers' understanding, such as layout, figures, or videos.

Future Directions

Future research should investigate the impact of nontext elements (ie, figures, infographics, videos) on information comprehension. In regions where two or more languages are spoken, it is also necessary to assess the readability of COVID-19 information in other languages. and differences in readability between languages should be assessed.

Conclusions

Readability levels of COVID-19 web-based information provided by the national COVID-19 portal of South Korea exceeded the recommended ninth-grade level. Efforts are needed to provide easy-to-read information to reach more people during a public health crisis. We hope that this study serves as a call to action for health care authorities to develop better guidelines that encourage an easy-to-read format so that information is provided at a level that most readers can understand and apply.

Authors' Contributions

HM and GHL designed the study and collected data. HM and YJC conducted statistical analyses. HM wrote the manuscript. All authors reviewed the final manuscript.

References

1. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
2. The second Korean adult literacy survey. National Institute for Lifelong Education.: The Korean Ministry of Education; 2017. URL: https://www.le.or.kr/contentView.do [accessed 2021-04-10]
3. Mark K, Elizabeth G, Ying J, Christine P. The health literacy of america's adults: results from the 2003 national assessment of adult literacy. National Center for Education Statistics. 2006. URL: https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006483 [accessed 2021-04-06]
4. History of health literacy definitions. Office of Disease Prevention and Health Promotion. URL: https://health.gov/our-work/healthy-people/healthy-people-2030/health-literacy-healthy-people-2030/history-health-literacy-definitions [accessed 2021-04-11]
5. Liu C, Wang D, Liu C, Jiang J, Wang X, Chen H, et al. What is the meaning of health literacy? a systematic review and qualitative synthesis. Fam Med Community Health 2020 Apr 14;8(2):e000351 [FREE Full text] [doi: 10.1136/fmch-2020-000351] [Medline: 32414834]
6. Dewalt DA, Berkman ND, Sheridan S, Lohr KN, Pignone MP. Literacy and health outcomes: a systematic review of the literature. J Gen Intern Med 2004 Dec;19(12):1228-1239 [FREE Full text] [doi: 10.1111/j.1525-1497.2004.40153.x] [Medline: 15610334]
7. Do BN, Tran TV, Phan DT, Nguyen HC, Nguyen TTP, Nguyen HC, et al. Health literacy, ehealth literacy, adherence to infection prevention and control procedures, lifestyle changes, and suspected COVID-19 symptoms among health care workers during lockdown: online survey. J Med Internet Res 2020 Nov 12;22(11):e22894 [FREE Full text] [doi: 10.2196/22894] [Medline: 33122164]
8. Huhta A, Hirvonen N, Huotari M. Health literacy in web-based health information environments: systematic review of concepts, definitions, and operationalization for measurement. J Med Internet Res 2018 Dec 19;20(12):e10273 [FREE Full text] [doi: 10.2196/10273] [Medline: 30567690]
9. Health literacy in healthy people 2030. Office of Disease Prevention and Health Promotion. URL: https://health.gov/our-work/healthy-people/healthy-people-2030/health-literacy-healthy-people-2030 [accessed 2021-04-15]
10. National action plan to improve health literacy. Office of Disease Prevention and Health Promotion. URL: https://health.gov/our-work/health-literacy/national-action-plan-improve-health-literacy [accessed 2021-04-07]
11. The 5th National Health Plan (HP2030). Korea Health Promotion Institute. URL: https://www.khealth.or.kr/healthplaneng [accessed 2021-04-12]

XSL•FO

RenderX

12.   van Deursen AJ. Digital inequality during a pandemic: quantitative study of differences in COVID-19-related internet uses and outcomes among the general population. J Med Internet Res 2020 Aug 20;22(8):e20073 [FREE Full text] [doi: 10.2196/20073] [Medline: 32750005]

13.   Brewer LC, Fortuna KL, Jones C, Walker R, Hayes SN, Patten CA, et al. Back to the future: achieving health equity through health informatics and digital health. JMIR Mhealth Uhealth 2020 Jan 14;8(1):e14512 [FREE Full text] [doi: 10.2196/14512] [Medline: 31934874]

14.   Laplante-Lévesque A, Thorén ES. Readability of internet information on hearing: systematic literature review. Am J Audiol 2015 Sep;24(3):284-288. [doi: 10.1044/2015_AJA-14-0091] [Medline: 26649531]

15.   Daraz L, Morrow AS, Ponce OJ, Farah W, Katabi A, Majzoub A, et al. Readability of online health information: a meta-narrative systematic review. Am J Med Qual 2018;33(5):487-492. [doi: 10.1177/1062860617751639] [Medline: 29345143]

16.   Walsh TM, Volsko TA. Readability assessment of internet-based consumer health information. Respir Care 2008 Oct;53(10):1310-1315 [FREE Full text] [Medline: 18811992]

17.   Weiss BD. Health Literacy and Patient Safety: Help Patients Understand. Manual for Clinicians. 2nd ed. Chicago: American Medical Association Foundation; Sep 05, 2007.

18.   Hyun-Cheol K. Achievements and prospects of easy-to-understand legislation project. Journal of Legislatioin Research Korea Legislation Research Institute 2019:51-79 [FREE Full text] [doi: 10.22851/kjlr.2019..57.003]

19.   Soo-Jin K, Tae-Wha L, Gwang-Suk K, Ju-Hee L. The levels of health literacy and related factors among middle-aged adults in Seoul, Korea. Korean Journal of Health Education and Promotion 2012 Sep 30:75-89 [FREE Full text] [doi: 10.1016/b978-1-4160-0016-7.50418-5]

20.   Koeze E, Popper N. The virus changed the way we internet. The New York Times. 2020 Apr 07. URL: https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html [accessed 2021-03-10]

21.   Husnayain A, Shim E, Fuad A, Su EC. Understanding the community risk perceptions of the COVID-19 outbreak in South Korea: infodemiology study. J Med Internet Res 2020 Sep 29;22(9):e19788 [FREE Full text] [doi: 10.2196/19788] [Medline: 32931446]

22.   Google's most-searched words of 2020. Google Trends. URL: https://trends.google.com/trends/yis/2020/GLOBAL/ [accessed 2021-03-31]

23.   Park S, Han S, Kim J, Molaie MM, Vu HD, Singh K, et al. COVID-19 discourse on twitter in four Asian countries: case study of risk communication. J Med Internet Res 2021 Mar 16;23(3):e23272 [FREE Full text] [doi: 10.2196/23272] [Medline: 33684054]

24.   Risk communications. World Health Organization. URL: https://www.who.int/emergencies/risk-communications [accessed 2021-04-03]

25.   United Nations Division for Public Institutions and Digital Government. UN/DESA Policy Brief #61: COVID-19: Embracing digital government during the pandemic and beyond. Department of Economic and Social Affairs. 2020 Apr 14. URL: https://tinyurl.com/2p8pajmv [accessed 2021-03-31]

26.   Bernstein C. What is digital health (digital healthcare) and why is it important? SearchHealthIT. URL: https://searchhealthit.techtarget.com/definition/digital-health-digital-healthcare [accessed 2022-02-15]

27.   Choi I, Yong C, Lee I. Readability of patient information leaflets in clinical trials. Korean J Clin Pharm 2016;26:33-39 [FREE Full text]

28.   Kim JE, Yang SJ. Evaluation of the readability and suitability of printed educational materials on metabolic syndrome. J Korean Pub Health Nurs 2016 Apr 30;30(1):149-163 [FREE Full text] [doi: 10.5932/JKPHN.2016.30.1.149]

29.   Lee TW, Kang SJ, Kim HH, Woo SR, Kim S. Suitability and readability assessment of printed educational materials on hypertension. J Korean Acad Nurs 2011 Jun;41(3):333-343. [doi: 10.4040/jkan.2011.41.3.333] [Medline: 21804342]

30.   Chin Y, Choi K. Readability and suitability evaluation of educational materials on diabetes mellitus. Korean J Health Serv Manag 2014 Jun 30;8(2):161-174 [FREE Full text] [doi: 10.12811/kshsm.2014.8.2.161]

31.   Szmuda T, Özdemir C, Ali S, Singh A, Syed M, Słoniewski P. Readability of online patient education material for the novel coronavirus disease (COVID-19): a cross-sectional health literacy study. Public Health 2020 Aug;185:21-25 [FREE Full text] [doi: 10.1016/j.puhe.2020.05.041] [Medline: 32516624]

32.   Worrall AP, Connolly MJ, O'Neill A, O'Doherty M, Thornton KP, McNally C, et al. Readability of online COVID-19 health information: a comparison between four English speaking countries. BMC Public Health 2020 Nov 13;20(1):1635 [FREE Full text] [doi: 10.1186/s12889-020-09710-5] [Medline: 33183297]

33.   Kruse J, Toledo P, Belton TB, Testani EJ, Evans CT, Grobman WA, et al. Readability, content, and quality of COVID-19 patient education materials from academic medical centers in the United States. Am J Infect Control 2021 Jun;49(6):690-693 [FREE Full text] [doi: 10.1016/j.ajic.2020.11.023] [Medline: 33259825]

34.   Mishra V, Dexter JP. Comparison of readability of official public health information about COVID-19 on websites of international agencies and the governments of 15 countries. JAMA Netw Open 2020 Aug 03;3(8):e2018033 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.18033] [Medline: 32809028]

35.   The national COVID-19 portal of South Korea. Korea Disease Control and Prevention Agency. URL: http://ncov.mohw.go.kr/ [accessed 2021-04-14]

36.   Park A, Zhu S, Conway M. The readability of electronic cigarette health information and advice: a quantitative analysis of web-based information. JMIR Public Health Surveill 2017 Jan 06;3(1):e1 [FREE Full text] [doi: 10.2196/publichealth.6687] [Medline: 28062390]

37.   Park SW. A Study on Utilization of Read-LQ. The Academy for Korean Language Education 2011 Apr;87(10):37-58 [FREE Full text] [doi: 10.15734/koed..87.201104.37]

38.   Kim G. Vocabulary for Korean language education by grade. In: Quantitative Linguistics 2nd edition. Hanam-si, Gyeonggi-do, Republic of Korea: Pagijong Press; Jan 20, 2003:157-356.

39.   What are the top 200 most spoken languages? Ethnologue: Languages of the World. Dallas, Texas: SIL International; 2021. URL: https://www.ethnologue.com/guides/ethnologue200 [accessed 2021-03-31]

40.   Key results of the statistics of the elderly. Statistics Korea. Daejeon, Republic of Korea; 2020 Sep. URL: http://kostat.go.kr/portal/korea/kor_nw/1/1/index.board?bmode=read&aSeq=385322 [accessed 2021-03-30]

41.   Perrotta F, Corbi G, Mazzeo G, Boccia M, Aronne L, D'Agnano V, et al. COVID-19 and the elderly: insights into pathogenesis and clinical decision-making. Aging Clin Exp Res 2020 Aug;32(8):1599-1608 [FREE Full text] [doi: 10.1007/s40520-020-01631-y] [Medline: 32557332]

42.   Kobayashi LC, Wardle J, Wolf MS, von Wagner C. Aging and functional health literacy: a systematic review and meta-analysis. J Gerontol B Psychol Sci Soc Sci 2014 Dec 11;71(3):445-457. [doi: 10.1093/geronb/gbu161] [Medline: 25504637]

43.   Lilleker D, Coman I, Gregor M, Novelli E. Governance and rhetoric in global comparative perspective. In: Political Communication and COVID-19. London: Routledge; Mar 19, 2021.

44.   Stanley C. Search Engine Market Share in Korea 2021. InterAd. 2020 Nov 13. URL: https://www.interad.com/en/category/insights/korean-search-engine-market-share.html [accessed 2020-12-30]

45.   Alexa top sites in South Korea. Alexa. URL: https://www.alexa.com [accessed 2022-02-15]

46.   Global top sites' ranking for health conditions and concerns. Similarweb. URL: https://www.similarweb.com/ [accessed 2022-02-15]

47.   Higgins O, Domegan C, Sixsmith J. A literature review on health information-seeking behaviour on the web?: a health consumer and health professional perspective. European Centre for Disease Prevention and Control. 2011 Oct 13. URL: https://data.europa.eu/doi/10.2900/5788 [accessed 2022-02-15]

48.   Healthcare workers: information on COVID-19. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/index.html [accessed 2021-03-28]

49.   Easy to Read COVID-19 Materials. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/easy-to-read/index.html [accessed 2021-03-28]

50.   Coronavirus guidance for clinicians and NHS managers. National Health Service England. URL: https://www.england.nhs.uk/coronavirus/ [accessed 2021-03-28]

51.   Coronavirus (COVID-19) (easy read). National Health Service Scotland. URL: https://www.nhsinform.scot/translations/formats/easy-read/coronavirus-covid-19-easy-read [accessed 2021-03-21]

52.   Valizadeh-Haghi S, Khazaal Y, Rahmatizadeh S. Health websites on COVID-19: are they readable and credible enough to help public self-care? J Med Libr Assoc 2021 Jan 01;109(1):75-83 [FREE Full text] [doi: 10.5195/jmla.2021.1020] [Medline: 33424467]

53.   Halboub E, Al-Ak'hali MS, Al-Mekhlafi HM, Alhajj MN. Quality and readability of web-based Arabic health information on COVID-19: an infodemiological study. BMC Public Health 2021 Jan 18;21(1):151 [FREE Full text] [doi: 10.1186/s12889-021-10218-9] [Medline: 33461516]

54.   Eltorai A, Sharma P, Wang J, Daniels A. Most American Academy of Orthopaedic Surgeons' online patient education material exceeds average patient reading level. Clin Orthop Relat Res 2015 Apr;473(4):1181-1186 [FREE Full text] [doi: 10.1007/s11999-014-4071-2] [Medline: 25475715]

55.   Rooney MK, Sachdev S, Byun J, Jagsi R, Golden DW. Readability of patient education materials in radiation oncology-are we improving? Pract Radiat Oncol 2019 Nov;9(6):435-440. [doi: 10.1016/j.prro.2019.06.005] [Medline: 31228657]

56.   Jayaratne YSN, Anderson NK, Zwahlen RA. Readability of websites containing information on dental implants. Clin Oral Implants Res 2014 Dec;25(12):1319-1324. [doi: 10.1111/clr.12285] [Medline: 24148074]

57.   Castro-Sánchez E, Spanoudakis E, Holmes AH. Readability of ebola information on websites of public health agencies, United States, United Kingdom, Canada, Australia, and Europe. Emerg Infect Dis 2015 Jul;21(7):1217-1219 [FREE Full text] [doi: 10.3201/eid2107.141829] [Medline: 26079313]

58.   Basch CH, Fera J, Garcia Bs P. Information regarding zika virus on the internet: a cross-sectional study of readability. Am J Infect Control 2020 Jun;48(6):714-715. [doi: 10.1016/j.ajic.2019.11.014] [Medline: 31862166]

59.   Emanuel EJ, Boyle CW. Assessment of length and readability of informed consent documents for COVID-19 vaccine trials. JAMA Netw Open 2021 Apr 01;4(4):e2110843 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.10843] [Medline: 33909052]

60.   Zhang M, Chow A, Smith H. COVID-19 contact-tracing apps: analysis of the readability of privacy policies. J Med Internet Res 2020 Dec 03;22(12):e21572 [FREE Full text] [doi: 10.2196/21572] [Medline: 33170798]

XSL•FO

RenderX

61.   Garcia P, Fera J, Mohlman J, Basch CH. Assessing the readability of covid-19 testing messages on the internet. J Community Health 2021 Oct;46(5):913-917 [FREE Full text] [doi: 10.1007/s10900-021-00973-6] [Medline: 33638806]

62.   Plain Language Materials & Resources. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/healthliteracy/developmaterials/plainlanguage.html [accessed 2021-03-31]

63.   Everyday Words for Public Health Communication. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/healthcommunication/everydaywords/ [accessed 2021-03-31]

64.   Plain language communication. Centers for Disease Control and Prevention. 2019 Oct 17. URL: https://www.cdc.gov/healthliteracy/developmaterials/plain-language-communication.html [accessed 2021-03-31]

65.   COVID-19 research update: how many pandemic papers have been published? Nature Index. 2020 Aug 28. URL: https://www.natureindex.com/news-blog/how-coronavirus-is-changing-research-practices-and-publishing [accessed 2022-02-15]

XSL•FO

RenderX

Original Paper

A Smartphone Serious Game for Adolescents (Grow It! App): Development, Feasibility, and Acceptance Study

Evelien Dietvorst[1*], MSc; Michelle A Aukes[1*], MSc; Jeroen S Legerstee[1], PhD; Annabel Vreeker[1], PhD; Micah M Hrehovcsik[2], PhD; Loes Keijsers[3], Prof Dr, PhD; Manon H J Hillegers[1], MD, Prof Dr, PhD

[1]Department of Child and Adolescent Psychiatry/Psychology, Erasmus MC Sophia Children's Hospital, Erasmus University Medical Center, Rotterdam, Netherlands

[2]Innovation Studio, HKU University of the Arts Utrecht, Utrecht, Netherlands

[3]Erasmus School of Social and Behavioural Sciences, Department of Psychology, Education & Child Studies, Rotterdam, Netherlands

[*]these authors contributed equally

Corresponding Author:
Manon H J Hillegers, MD, Prof Dr, PhD
Department of Child and Adolescent Psychiatry/Psychology
Erasmus MC Sophia Children's Hospital
Erasmus University Medical Center
Wytemaweg 8
Rotterdam, 3015 CN
Netherlands
Phone: 31 (0)107040209
Email: m.hillegers@erasmusmc.nl

Abstract

Background: Anxiety and mood problems in adolescents often go unnoticed and may therefore remain untreated. Identifying and preventing the development of emotional problems requires monitoring and effective tools to strengthen adolescents' resilience, for example, by enhancing coping skills.

Objective: This study describes the developmental process, feasibility, and acceptance of Grow It!, a multiplayer serious game app for adolescents aged 12-25 years. The app consists of the experience sampling method (ESM) to monitor thoughts, behaviors, and emotions in daily life to enhance self-insight and daily cognitive behavioral therapy–based challenges to promote adaptive coping.

Methods: Our approach entails an iterative game design process combined with an agile method to develop the smartphone app. The incorporated game features (ie, challenges, chat functionality, and visual representation) in the Grow It! app were co-designed with adolescent end users to increase participant engagement and adherence.

Results: The Grow It! app was delivered for Android and iOS in May 2020. Grow It! was offered to adolescents during the COVID-19 crisis between May and December 2020. Participants of the Grow It! COVID-19 study (sample 1: N=685; mean age 16.19, SD 3.11 years; 193/685, 28.2% boys; sample 2: N=1035; mean age 18.78, SD 3.51 years; 193/1035, 18.64% boys) completed 31.5% (13.2/42) to 49.5% (10.4/21) of challenges. Compliance of ESM was suboptimal (35.1/210, 16.7% to 32.5/105, 30.9%). Follow-up questionnaires indicated an overall score of the app of 7.1 out of 10. Moreover, 72.6% (278/383) to 75.6% (487/644) would recommend the app to friends.

Conclusions: To our knowledge, Grow It! is the first gamified ESM app that both measures individual differences in emotional dynamics and offers an integrated cognitive behavioral therapy–based intervention. Our findings support the feasibility and acceptance, and therefore applicability, of the Grow It! app in adolescents. Further iterations of this serious game app will focus on the increase of compliance and on providing participants feedback through their personal mood profiles.

Introduction

Background

Internalizing problems, such as anxiety and mood problems, have a substantial impact on young people's lives. These internalizing problems are often associated with school dropout, reduced social functioning, loneliness, unemployment, and reduced quality of life [1-3]. Anxiety and mood disorders usually begin in adolescence [4,5], and unfortunately, they often go unnoticed or untreated [6]. When persistent, internalizing problems often result in emerging psychiatric disorders affecting young people's daily lives, their future, and society [1-3,6]. Therefore, early identification and timely intervention are crucial to prevent further deterioration, improve prognosis, and reduce the burden on health care systems and society in general [1,7,8].

Mobile health (mHealth) can play an important role in accurate recognition of symptoms and timely treatment [9-11]. mHealth is defined as wireless technologies, such as smartphone apps, to support or achieve health objectives. In terms of its advantages, first of all, mHealth is scalable, accessible, and maybe less stigmatizing than traditional treatment for youths because of the level of anonymity and privacy [12]. Furthermore, mHealth offers the possibility of incorporating motivational elements such as playfulness and gamification, which is advantageous because humans supposedly learn best by playing [13-17]. Finally, mHealth offered in an attractive and fun way through adolescents' own devices fits very well with their daily life and activities [18,19]. Use is flexible, as it is independent of time and place and can be at a self-determined pace, which is thought to enhance self-efficacy [12]. Indeed, most adolescents indicated that they would use an app to screen for emotional problems and treatment if available [20]. Even though there are already several mHealth apps [21], preventive mHealth apps that integrate novel methods for early identification and preventive intervention are still lacking.

A promising method for the early identification of emotional problems is the experience sampling method (ESM). The ESM is a structured diary method in which participants obtain multiple random notifications on their phone during the day. When a notification pops up, they fill out a microquestionnaire in which they report on their behaviors, thoughts, and feelings in real time (eg, how are you doing right now?). The strength of the ESM is the high ecological validity. Moment-to-moment assessments in real-world settings address the problem of recall bias [22,23]. The promise of ESM data for early identification of adolescents at risk for the development of psychopathology has been demonstrated in a research setting with adults [24] and adolescents [25]. Moreover, self-management may be enhanced by obtaining insights into everyday functioning dynamics, based on ESM data [22,26,27].

The first choice of treatment of anxiety and mood problems in adolescents is psychological therapy, especially cognitive behavioral therapy (CBT) [28-30]. It is known that the way adolescents cope with stress or handle negative emotions in daily life may increase or buffer against the development of anxiety and depressive symptoms [31-33]. That is why CBT is one of the effective interventions aimed at improving coping. Recently, results of an internet-based CBT intervention revealed that positive effects occur already after 4 weeks of CBT [34]. CBT is mostly used in clinical practice but increasingly also applied for preventive purposes [28,29].

Therefore, we cocreated the Grow It! app (Android and iOS) for adolescents aged 12 to 25 years. Grow It! is a multiplayer serious game, a game that is designed for a primary purpose other than entertainment [30]. Incorporated in the Grow It! app are ESM, to enhance self-management and identify mood problems earlier on, and gamified CBT-based challenges, to increase coping. Initially, the app was developed for high-risk adolescent populations, such as adolescents with chronic somatic conditions, offspring of parents with psychiatric disorders, or adolescents experiencing extreme stressful societal circumstances, for example, the COVID-19 pandemic. However, the app may also serve a broader purpose of prevention for adolescents from a general population.

Objectives

The first aim of this study is to give an elaborate description of the developmental process of the multiplayer serious game app Grow It! Second, we aim to study the feasibility and acceptance of Grow It! among end users.

Methods

Developmental Process

Overview

The Grow It! app was developed, with intermittent periods, from March 2016 to February 2020 (Figure 1). During its development, we cocreated the app with a large multidisciplinary team of child and adolescent psychiatrists, developmental and clinical psychologists, data analysts, game designers, and multiple test panels (adolescents aged 12-25 years) [35,36]. The initial concept was developed with University of the Arts Utrecht (test 1a-1b). With IJsfontein BV and consultancy company Game Architect Studio, an agile process (defined as a software development methodology including iterative development, where requirements and solutions evolve in multidisciplinary teams [37]) then was used to develop and evaluate the minimal viable product (MVP; test 2a-2e). For all tests, informed consent was obtained.

Figure 1. The developmental process of the Grow It! app. ESM: experience sampling method; VAS: visual analog scale.

Initial Concept

In March 2016, the developmental process of Grow It! started with a prepilot (test 1a, N=10), which resulted in a paper-based prototype and wireframes. Later on, in the pilot (test 1b), adolescents from the general population (N=21) tested the app's beta version during a 6-week trial. Weeks 1 and 6 of the test consisted of an ESM-only study with 8 mood assessments per day. During weeks 2 to 5, the mood assessments were given twice a day and combined with CBT-based challenges. In all weeks, users received feedback through push messages complimenting them. Lessons learned were that adolescents were motivated by the game mechanics of Grow It! and liked completing the ESM questionnaires and daily challenges.

Minimal Viable Product

In June 2019, an MVP was built. We aimed to improve the app's content, visual design, interaction design, and reliability of assessments and ran a technical test. The app was developed using agile development and user-centered design methods, including different tests and collaboration with focus groups. Different groups of adolescents (n=6 and n=9) received instructions and were invited to design CBT-based challenges aiming at adaptive coping [36]. Thereafter, all ideas were formulated into specific challenges and were rated in terms of their clinical appropriateness and coping effectiveness by 11 child and youth psychologists and psychiatrists. As a result of these focus groups (test 2a), 126 challenges were formulated, which were later used as the challenges in the Grow It! app. Furthermore, a survey among 107 adolescents (test 2b) resulted in the choice for a visual design that (1) was accepted by a broad range of ages and by both boys and girls and (2) was low cost in maintenance. In interviews (N=4, test 2c) with regard to the navigation of the app, special attention was given to the answer-scale development (Likert and visual analog scale) to increase the assessments' reliability. In empirical studies on ESM data [22], visual analog scales had demonstrated the anchor-influenced results, and therefore, different approaches

were tested (eg, add a cursor above the anchor line). Interviews with adolescents indicated that Likert scales are more intuitive if they run from high to low on the screen instead of low to high. In December 2019 and February 2020, user experience tests of the Grow It! app were run (test 2d). Adolescents (N=23) played the Grow It! app for 1 week. On the basis of interviews, we improved and extended the content of the app by (1) adding a limited chat function with predesigned stickers to motivate team members, (2) allowing users to choose from 3 challenges a day for 6 weeks, and (3) adding a tour in which game mechanics are explained. Finally, a handbook for errors arose from our technical test and quality assessment, which was performed by the research team (N=10, test 2e).

Feasibility and Acceptance Test

In May 2020 and in December 2020, at the first 2 peaks of the COVID-19 pandemic, the Grow It! app was launched to assess the game mechanics and user acceptance of the MVP. Owing to government restrictions, all adolescents had to follow social distancing measures (eg, staying at home because schools were closed). Through (social) media, the app was made available to Dutch-speaking adolescents living in the Netherlands, aged 12-25 years, who owned a smartphone. Participants were consecutively enrolled in a Grow It! team after completion of the baseline questionnaire that was linked to the web-based informed consent procedure on a secure website. This way, participants started with the app as soon as possible.

In total, 685 adolescents (sample 1: mean age 16.19, SD 3.11 years; 193/685, 28.2% boys) played the Grow It! app for 6 weeks, and in the second sample, another 1035 adolescents (sample 2: mean age 18.78, SD 3.51 years; 193/1035, 18.64% boys) played the Grow It! app for 3 weeks. A follow-up questionnaire was filled out by 383 and 644 adolescents for samples 1 and 2, respectively (see Table 1 for demographics). In the Grow It! app, participants were given 5 ESM notifications per day and daily challenges. Users who did not show activity in the app (0 or 1 activity in ESM or challenges) were excluded

from the sample, because to evaluate the user experience of the Grow It! app we were interested in participants who were involved in playing the app.

A complete overview of all (ESM) instruments and questionnaires is provided in our internet-based codebook [38].

The outcomes of our feasibility and acceptance test can be found in the *Results* section. Statistical analyses are descriptive (means, SDs, and frequencies) and performed using SPSS (version 25; IBM Corp) [39].

Table 1. Sample characteristics and demographics.

	Sample 1		Sample 2	
	App engagement (Grow It! activity; N=685)	Follow-up questionnaire (n=383; 55.9% retention)	App engagement (Grow It! activity; N=1035)	Follow-up questionnaire (n=644; 62.2% retention)
Age (years), mean (SD)	16.19 (3.11)	16.26 (3.07)	18.78 (3.51)	18.48 (3.43)
Gender, n (% boys)	193 (28.2)	100 (26)	193 (18.6)	120 (18.7)
Education level[a], n (%)				
Primary school	30 (4.4)	9 (2.3)	9 (0.9)	6 (1)
Low	104 (15.2)	54 (14.1)	167 (16.1)	98 (15.2)
Medium	152 (22.1)	92 (24.1)	337 (32.6)	201 (31.2)
High	399 (58.2)	228 (59.5)	438 (42.3)	293 (45.5)
Other	N/A[b]	N/A	84 (8.1)	46 (7.1)
Cultural identity, n (%)				
Dutch	622 (90.8)	348 (90.7)	1013 (97.9)	631 (98.1)
Mixed	57 (8.3)	4 (1.1)	17 (1.6)	11 (1.7)
Other	6 (0.9)	31 (8.1)	5 (0.6)	2 (0.3)

[a]Low: (preparatory school for) technical and vocational training; medium: (preparatory school for) professional education; and high: (preparatory school for) university.

[b]N/A: not applicable.

End Product of Developmental Process: The Grow It! App

User Journey

The user journey first entails a phase of enrollment, during which participants can personalize their account. After receiving a 6-digit code (letters and numbers) from the research team via SMS text messaging, they log in to Grow It! and choose their nickname based on 2 turntables. The first turntable shows an adjective (eg, Adorable, Dangerous, Lucky, Creative, and Romantic), and the second turntable shows an animal name (eg, Alpaca, Snake, Iguana, Rabbit, and Crocodile). Participants can rotate the turntables as often as they want to personalize their nickname. For example, one participant nickname could be *Lucky Rabbit* or *Adorable Alpaca* (Figure 2). The game mechanics (ie, personalization, collaboration, competition, and feedback) are explained in the mandatory tour of the Grow It! app.

Figure 2. Nickname.
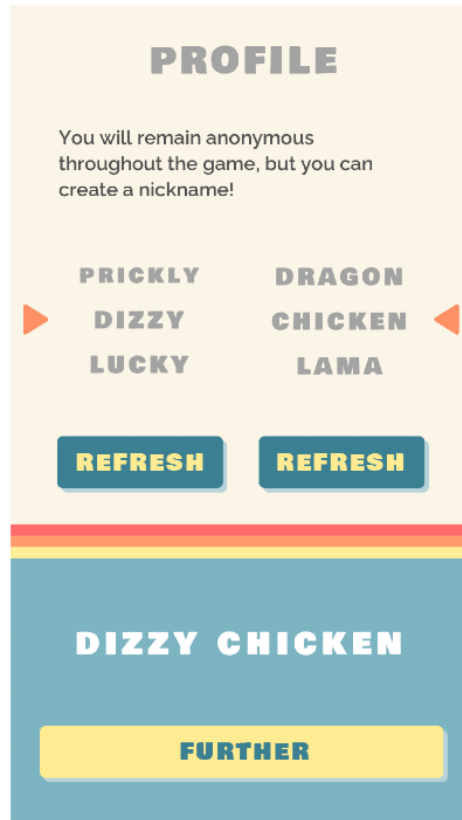
Collaboration

As adolescents are sensitive to peer influence and can be motivated by interactions with peers [40-42], each participant collaborates anonymously in a team with 3 to 7 other players.

Adolescents are allocated to a team by the researchers. To support team members, participants can chat by sending and receiving positive stickers (Figure 3). Via this chat system, participants can motivate each other, while the system minimizes the possibilities of bullying and negative peer pressure.

Figure 3. Chat function with positive stickers.
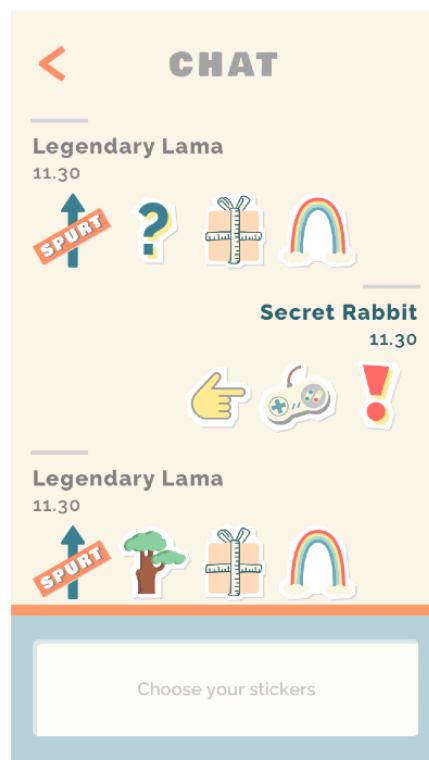
Competition

Competition is encouraged at the team level, where teams play versus each other. Each team has a virtual tree with a name (eg, Oaks, Pines, and Palms), which allows participants to compare their team performance with that of other teams (Figure 4). At the start of the game, the tree is empty. Teams *grow* their tree when participants receive points by reporting their feelings and behaviors (ESM) and doing daily challenges. These reports are personal and shielded. Team members only see the amount of points their teammates have collected. After a team collects a specified amount of points, they achieve a *spurt* (ie, level-up), which means that the tree grows in height, and every team member receives a gift to embellish the tree (Figures 5 and 6). As game mechanics, these provide a positive feedback loop and a progress update and establish the reward scheme or the

behavioral conditioning that increases retention. Upon earning a gift (ie, loot box) from a growth spurt, a participant can then select his or her choice from 3 gifts. The gifts are wrapped so that there is no indication of what is inside. As a game mechanic, selecting a random gift creates surprise and moments of anticipation essential to maintaining a state of play [43]. When teams have just started using the app, it does not take many points to achieve the first *spurt* and earn a gift. As the game progresses, however, and teams move to higher levels, more and more points are needed to earn a *spurt* and gifts. In this way, adolescents are stimulated to keep playing and remain engaged with the app. The difficulty level is scaffolded by incrementally increasing difficulty, which supports retention by continuously challenging the participants as they progress through the game.

Figure 4. Comparing own tree with trees of other teams.
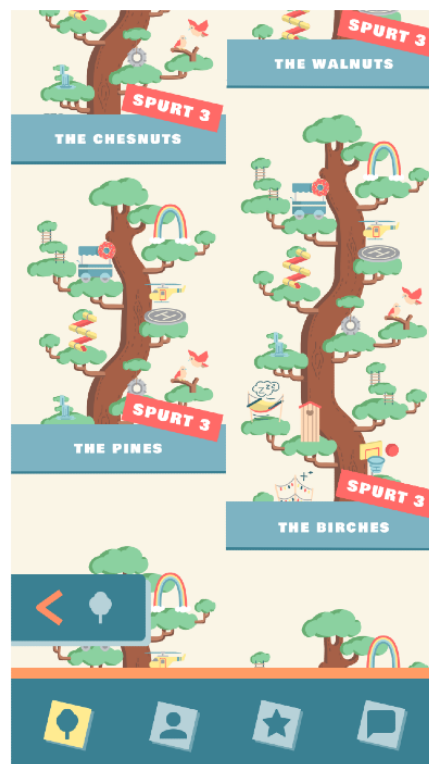
Figure 5. Receiving a gift.

Figure 6. Tree decorated with gifts.

Feedback

The game mechanics of Grow It! provide feedback at different levels. Whereas users can see their own and their team members' scores in the score overview screen (Figure 7), on their profile page, they obtain an overview of how many times they reported their feelings, behaviors, and challenges that day (Figure 8). Finally, the Grow It! app has a contact button in case of technical issues or urgent psychological problems. Whenever a participant pushes the contact button, a phone number is displayed through which the research team can be reached by telephone or texting on working days during office hours. On the study website, information can be found on how to reach help in acute situations or outside working hours, referring to professional and free services.

Figure 7. Overview of earned points.
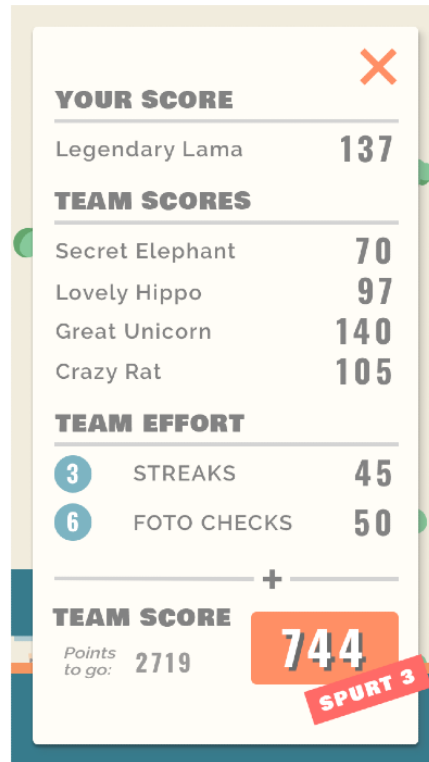
Figure 8. Profile page with the number of completed questionnaires.
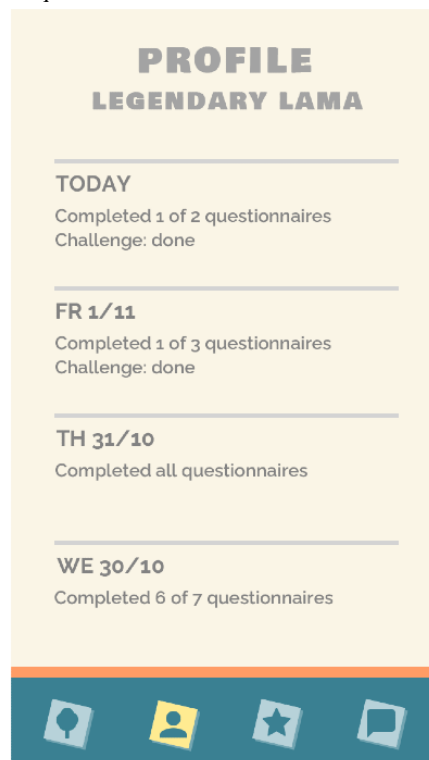
Daily Emotions

The ESM is an integral part of the app, which can be used for early identification of emotional problems and enhancing self-management in adolescents [22]. To prompt adolescents to report on their feelings, they receive several notifications per day, which are randomized to prevent structural answering patterns (eg, always in math class at 11 AM) [22]. In the first studies, adolescents answered 5 microquestionnaires per day (taking approximately 1-2 minutes) regarding their sleep, activity, affective well-being (eg, I feel happy or sad), coping strategies, pain, fatigue, social behavior, loneliness, stress, and coping (Figure 9).

Figure 9. Examples of daily experience sampling method questions.

Our ESM approach's novelty is that it is gamified to increase motivation, which intends to result in a higher rate of compliance (percentage completed self-evaluations) as well as improved data quality [15]. Compliance is one of the critical quality markers for ESM studies [22].

Daily Challenges

To teach adolescents how to cope with setbacks and to promote emotional resilience, the Grow It! app contains daily challenges aimed at strengthening adaptive coping, supporting physical activation, and preventing emotional problems. Coping styles incorporated in the challenges promote distraction, problem solving, social support, and acceptance [44,45]. Participants can choose 1 out of 3 challenges per day (Figure 10). Challenges are divided into three categories: photo challenges, quizzes, and assignments. Examples of challenges are as follows: make a picture of something you hold dear (photo challenge; aimed at distraction), ask someone what they like about you and write it down (assignment; aimed at social support), or answer a multiple-choice question such as *What is sushi usually rolled in?* (quiz; aimed at problem solving). An additional randomly available assignment is the photo check. Participants are shown a matrix of 9 photos and assess which photos fit a particular theme. In this way, they act as the photo challenge jury and can award points to participants and earn extra points themselves for this task.

Figure 10. Daily challenges.

Ethical Considerations and Privacy

Risks related to privacy were mitigated by making all participants pseudonymous and only identifiable in the app by participation codes and pseudonyms. Participants determine their pseudonym (ie, nickname) from a preselected set of words provided by the app. Participants cannot be identified, and only the research team has insight into the private data of the participants. Data collected with the app pertain to the user's game data (eg, game-specific actions) and responses to the ESM. The app also accesses the mobile device's camera, but only when a participant takes a photo for a challenge (not at other times), and the app does not access other functionalities (eg, Google, GPS, or health apps). All user data are encrypted and sent directly to a secure server at the researcher's institute. The privacy and security of the Grow It! app is approved by the privacy and security office of Erasmus Medical Center, and the app complies with the Dutch General Data Protection Regulation (Algemene Verordening Gegevensbescherming) and NEN-norm 7510:2017 (Dutch standard of information security management systems in health care). The app is available for research purposes in the Google Play store [46] and Apple store [47]. This study has been approved by the Medical Ethics Committee of the Erasmus Medical Center (MEC-2020-0287).

Results

Overview

As shown in Table 1, participants of sample 1 (N=685; mean age 16.2, SD 3.1 years) were somewhat younger than adolescents in sample 2 (N=1035; mean age 18.8, SD 3.5 years). In general, participants of both samples were relatively highly educated and mostly of Dutch ethnicity. The follow-up questionnaire was filled out by 55.9% (383/685) and 62.2% (644/1035) of the users in sample 1 and sample 2, respectively.

Feasibility and Acceptance

An overview of app engagement (ESM compliance and challenges) and answers of the user evaluation questionnaire carried out in the follow-up can be found in Tables 2 and 3.

Regarding the ESM component, overall compliance was 16.7% (35.1/210 notifications, sample 1) and 30.9% (32.5/105 notifications, sample 2). About 56.77% (583/1027) of the participants indicated that they thought the number of questions per day was too high, whereas about 40.7% (418/1027) indicated the number of questions as sufficient. Some participants also reported that they did not understand why the same questions were asked repeatedly (sample 1: N=20, sample 2: N=16; ie, at each notification, the same ESM questions were asked to monitor their feelings and behavior over the study weeks). Moreover, 15% (97/644) to 22.9% (88/383) reported no effect of the ESM. However, 66.8% (256/383) to 72.4% (466/644) reported reflecting more on their feelings as a result of the ESM.

With regard to the daily CBT-based challenges, participants completed 31.5% (216/685) to 49.47% (512/1035) of all challenges. Whereas in sample 1, a total of 1.2% (8/685) participants completed all 42 challenges (100%), 6.57% (68/1035) participants completed all 21 challenges (21/21, 100%) in sample 2. The self-reported effect of the challenges showed that 20.6% (79/383) to 44.2% (285/644) of the participants became more active as a result of the Grow It! challenges.

The overall user evaluation of Grow It! was positive. The average app evaluation score was 7.1 of 10 (SD 1.5) in sample 1 and 7.2 of 10 (SD 1.3) in sample 2. Moreover, the app's design was evaluated with a score of 7.7 to 8.0 of 10. Finally, 72.6% (278/383) to 75.6% (487/644) would recommend the app to their friends.

Table 2. App engagement and user evaluation of Grow It! (sample 1).

	Total	Aged 12-17 years	Aged 18-25 years	Difference test (aged 12-17 vs 18-25 years)		
				t test (df)	Chi-square (df)	P value
App engagement (Grow It! activity)					N/A[a]	
Number of users	685	500	185	N/A		N/A
Compliance of ESM[b] (n=210), number of notifications (%)	35.1 (16.7)	34.4 (16.4)	41.8 (19.9)	1.89 (683)		.06
Challenges (n=42), n (%)	13.2 (31.5)	13.3 (31.6)	13.1 (31.2)	0.19 (683)		.85
User evaluation (follow-up questionnaire)					N/A	
Number of users	383	273	110	N/A		N/A
Evaluation of the app (1-10), mean (SD)	7.1 (1.5)	7.4 (1.3)	6.6 (1.7)	4.70 (381)		<.001
Evaluation of the design (1-10), mean (SD)	7.7 (1.5)	7.8 (1.5)	7.5 (1.5)	2.21 (381)		.03
Self-reported effect of ESM, n (%)				N/A	1.99 (1)	.58
I got to know myself better	20 (5.3)	17 (6.3)	3 (2.9)			
It made me feel better	19 (5)	13 (4.7)	6 (5.7)			
It made me think about how I feel more	256 (66.8)	181 (66.4)	75 (67.6)			
No effect	88 (22.9)	62 (22.5)	26 (23.8)			
Evaluation amount of ESM per day, n (%)				N/A	9.18 (1)	.002
Few	15 (4)	8 (3)	7 (6.5)			
Sufficient	155 (40.3)	122 (44.8)	32 (29)			
A lot	213 (55.7)	143 (52.2)	71 (64.5)			
Self-reported effect of challenges, n (%)				N/A	5.26 (1)	.26
I have had more contact with others	10 (2.5)	6 (2.3)	3 (2.9)			
I am better at solving problems	8 (2.2)	4 (1.5)	4 (3.8)			
I am better at accepting situations	32 (8.3)	22 (8.1)	10 (8.7)			
I have become more active	79 (20.7)	64 (23.6)	15 (13.5)			
I have started to feel less lonely	28 (7.4)	20 (7.3)	8 (7.7)			
Evaluation chat function, n (%)				N/A	5.79 (1)	.22
Not nice at all	79 (20.7)	51 (18.7)	28 (25.5)			
A little bit nice	175 (45.7)	122 (44.7)	53 (48.1)			
Quite nice	54 (14.1)	41 (14.9)	14 (13.3)			
Nice	45 (11.7)	33 (12.2)	11 (10.4)			
Very nice	30 (7.9)	26 (9.5)	4 (3.8)			
Would recommend Grow It! to friends, n (%)	278 (72.6)	212 (77.7)	66 (60)	N/A	12.28 (1)	<.001
Want to help with future development of the app, n (%)	133 (34.7)	93 (33.9)	41 (36.9)	N/A	0.35 (1)	.55

[a]N/A: not applicable.

[b]ESM: experience sampling method.

Table 3. App engagement and user evaluation of Grow It! (sample 2).

	Total	Aged 12-17 years	Aged 18-25 years	Difference test (aged 12-17 vs 18-25 years)		
				t test (df)	Chi-square (df)	P value
App engagement (Grow It! activity)					N/A[a]	
Number of users	1035	405	630	N/A		N/A
Compliance of ESM[b] (n=105), number of notifications (%)	32.5 (30.9)	30 (28.6)	33.9 (32.3)	2.17 (1033)		.03
Challenges (n=21), n (%)	10.4 (49.5)	10.5 (50.1)	10.3 (49.2)	0.45 (1033)		.65
User evaluation (follow-up questionnaire)					N/A	
Number of users	644	256	388	N/A		N/A
Evaluation of the app (1-10), mean (SD)	7.2 (1.3)	7.5 (1.3)	6.9 (1.3)	1.59 (624)		.11
Evaluation of the design (1-10), mean (SD)	8.0 (1.3)	8.2 (1.3)	7.8 (1.3)	4.48 (624)		<.001
Self-reported effect of ESM, n (%)				N/A	8.54 (1)	.003
I got to know myself better	62 (9.6)	27 (10.4)	35 (9)			
It made me feel better	19 (3)	13 (5.2)	6 (1.5)			
It made me think about how I feel more	466 (72.4)	174 (68.0)	292 (75.3)			
No effect	97 (15)	42 (16.4)	55 (14.2)			
Evaluation amount of ESM per day, n (%)				N/A	5.77 (1)	.06
Few	11 (1.7)	5 (1.8)	7 (1.7)			
Sufficient	263 (40.9)	119 (46.5)	144 (37.2)			
A lot	370 (57.4)	132 (51.6)	237 (61.2)			
Self-reported effect of challenges, n (%)				N/A	10.47 (1)	.001
I have had more contact with others	106 (16.5)	31 (12)	78 (20)			
I am better at solving problems	29 (4.5)	16 (6.3)	12 (3)			
I am better at accepting situations	120 (18.7)	47 (18.3)	74 (19)			
I have become more active	285 (44.2)	121 (47.2)	162 (41.7)			
I have started to feel less lonely	104 (16.1)	41 (16.2)	62 (16.1)			
Evaluation chat function, n (%)				N/A	28.11 (1)	<.001
Not nice at all	142 (22)	35 (13.5)	107 (27.6)			
A little bit nice	288 (44.7)	109 (42.7)	179 (46)			
Quite nice	106 (16.4)	54 (21.2)	51 (13.2)			
Nice	77 (12)	41 (16.1)	36 (9.4)			
Very nice	31 (4.9)	17 (6.6)	15 (3.8)			
Would recommend Grow It! To friends, n (%)	487 (75.6)	215 (84)	272 (70)	N/A	16.12 (1)	<.001
Want to help with future development of the app, n (%)	263 (40.9)	117 (45.7)	146 (37.7)	N/A	0.74 (1)	.39

[a]N/A: not applicable.

[b]ESM: experience sampling method.

Results for Adolescents Compared With Emerging Adults

Given the broad age range in which adolescents participated (12-25 years), we also reported outcomes separately for adolescents (12-17 years) and emerging adults (18-25 years) in Tables 2 and 3. Higher user evaluations of the Grow It! app were found in the adolescent group in comparison with the emerging adult group (ie, samples 1 and 2: adolescents rated the design of the app higher, and more adolescents would recommend the app to their friends in samples 1 and 2; sample 1: adolescents' evaluation of the app was higher, and number of ESM notifications per day was evaluated better; sample 2: evaluation of the chat function was better in adolescents). Moreover, no differences between age groups were found with regard to app engagement with the exception of slightly higher

compliance of ESM in emerging adults than in adolescents (only in sample 1).

Discussion

Principal Findings

The Grow It! app emerged from a lack of and need for preventive interventions for emotional problems and promoting adaptive coping for adolescents; in addition, these interventions need to be low key, nonstigmatizing, fun, attractive, private, and secure. On the basis of the developmental process and acceptance and feasibility of Grow It!, key lessons learned and directions for future research are formulated and shared. Our approach entails an iterative game design process combined with an agile method to develop the smartphone app. The incorporated game features in the Grow It! app were co-designed with adolescent end users to increase participant engagement and adherence. With regard to the app engagement and user evaluation filled out at follow-up, we indicated that we have some evidence that supports the feasibility and acceptance, and therefore applicability, of Grow It! in adolescents.

Earlier studies suggest that adolescents are open to using mHealth [48]. Indeed, the large interest as well as the positive user evaluation provide reason to believe that adolescents, at least a large group, are positive and open to using mHealth. Grow It was co-designed with youths, which was reflected in age-adequate daily challenges and an overall positive rating of the Grow It! app. Moreover, 66.8% (256/383) to 72.4% (466/644) of the participants felt they reflected more on their emotions.

Limitations

Although the results of the developmental process, acceptance, and feasibility are informative and promising for mHealth, several limitations should be mentioned. First, the majority of the study sample consisted of girls. The design of the app seems more appealing to girls, despite the long process of cocreation with both boys and girls. One of the explanations might be that girls are more inclined to seek help [49] and therefore are also more inclined to participate in our study.

Second, the COVID-19 pandemic and related governmental restrictions might have influenced the results. Owing to remote working during the COVID-19 pandemic, participants had no or minimal personal contact and received minimal instructions on how to participate in an ESM study, although this is stated as an important factor to obtain reliable data [22]. Third, concerning self-monitoring, it is notoriously difficult to motivate adolescents. In this study, compliance of ESM was lower than in a typical research design [22]; however, participants received no financial incentive and were entirely motivated by the game structure to answer questionnaires. We, therefore, expect that with more targeted use (eg, in blended care when the app is explained by a professional), the compliance and user satisfaction might be more favorable. Furthermore, the dropout rates were comparable with other web-based studies [50]. In order to increase app engagement, improvements are needed; for instance, designs may need to be tailored to the individual (personalization) and (in-person) feedback is needed.

Future Directions

With regard to the early identification of emotional problems, ESM data of the Grow It! app provides an opportunity to develop algorithms in future research for the early detection of emotional problems, which often go unnoticed in adolescents [6]. Identifying emotional problems early on would require capitalizing on novel developments in clinical psychology [51] combined with the motivational game architecture codeveloped with youths [36]. The ESM has already shown to be a reliable method to investigate variation in thoughts, feelings, and symptoms over time and context in research settings [22,23]. Dietvorst et al [25] have demonstrated that ESM data helps to identify the onset of depressive feelings among adolescents 3 months ahead. Specifically, this was done by differentiating typical adolescents (eg, grumpy at home), from early depressive feelings. In future work, analyzing highly rich ESM data with more powerful analytical techniques, such as machine learning, could potentially improve this early identification.

In clinical practice, self-management and self-insight may be enhanced by obtaining insights into one's emotion dynamics [22,26,27]. A feature such as providing participants with feedback through a daily life emotion chart (eg, mood profile) could provide participants with better insight and feedback into their well-being [52,53]. It may also serve as a therapeutic function, as integrating real time mood profiles in the app could encourage adolescents to reflect more on their emotions, coping, and behavior in different contexts [26] and could also be used as a routine outcome measure during therapy. To test the effects of the app upon adolescent well-being and resilience, an additional in-depth evaluation is required. Research questions and hypotheses that are beyond the scope of the developmental process focusing on the main effect of the app are preregistered [54,55] and will be executed in the future accordingly, including a randomized controlled trial study to test the effectiveness of the Grow It! app.

Conclusions

The Grow It! app has been developed and improved through iterations in collaboration with a large multidisciplinary team. It is innovative, age-attuned, easily accessible, fun, and visually appealing and, most importantly, serves the needs of adolescents. The app was well received by adolescents, and the first findings presented here indicate that adolescents were motivated by the game mechanics of Grow It! and liked completing the ESM questionnaires and daily challenges. Initially, the app was developed for high-risk adolescent populations, such as adolescents with chronic somatic conditions, offspring of parents with psychiatric disorders, or adolescents experiencing extreme stressful societal circumstances, for example, the COVID-19 pandemic. However, the app may also serve a broader purpose of prevention for adolescents from a general population. Our findings support the feasibility and acceptance, and therefore applicability, of the Grow It! app in adolescents.

The ambition is to further improve the app after each research study by including new features and resolving usability issues. The next step will be to focus on the increase of compliance and providing participants feedback through their personal mood profiles.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Thapar A, Collishaw S, Pine DS, Thapar AK. Depression in adolescence. The Lancet 2012 Mar;379(9820):1056-1067. [doi: 10.1016/S0140-6736(11)60871-4]

2. Cummings CM, Caporino NE, Kendall PC. Comorbidity of anxiety and depression in children and adolescents: 20 years after. Psychol Bull 2014 May;140(3):816-845 [FREE Full text] [doi: 10.1037/a0034733] [Medline: 24219155]

3. Adolescent mental health. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health [accessed 2022-01-27]

4. Garber J, Keiley MK, Martin NC. Developmental trajectories of adolescents' depressive symptoms: predictors of change. J Consult Clin Psychol 2002 Feb;70(1):79-95. [doi: 10.1037//0022-006x.70.1.79] [Medline: 11860059]

5. Ormel J, Raven D, van Oort F, Hartman CA, Reijneveld SA, Veenstra R, et al. Mental health in Dutch adolescents: a TRAILS report on prevalence, severity, age of onset, continuity and co-morbidity of DSM disorders. Psychol Med 2015 Jan;45(2):345-360. [doi: 10.1017/S0033291714001469] [Medline: 25066533]

6. Kessler RC, Angermeyer M, Anthony JC, Graaf RDE, Demyttenaere K, Gasquet I, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry 2007 Oct;6(3):168-176. [Medline: 18188442]

7. Leijdesdorff SM, Huijs CE, Klaassen RM, Popma A, van Amelsvoort TA, Evers SM. Burden of mental health problems: quality of life and cost-of-illness in youth consulting Dutch walk-in youth health centres. J Ment Health 2020 Oct 22:1-8. [doi: 10.1080/09638237.2020.1836555] [Medline: 33086874]

8. Raven D, Jörg F, Visser E, Oldehinkel AJ, Schoevers RA. Time-to-treatment of mental disorders in a community sample of Dutch adolescents. A TRAILS study. Epidemiol Psychiatr Sci 2017 Apr;26(2):177-188 [FREE Full text] [doi: 10.1017/S2045796016000226] [Medline: 27075651]

9. Eijlers R, Legerstee JS, Dierckx B, Staals LM, Berghmans J, van der Schroeff MP, et al. Development of a virtual reality exposure tool as psychological preparation for elective pediatric day care surgery: methodological approach for a randomized controlled trial. JMIR Res Protoc 2017 Sep 11;6(9):e174 [FREE Full text] [doi: 10.2196/resprot.7617] [Medline: 28893727]

10. Granic I, Lobel A, Engels RC. The benefits of playing video games. Am Psychol 2014 Jan;69(1):66-78. [doi: 10.1037/a0034857] [Medline: 24295515]

11. Watts SE, Andrews G. Internet access is NOT restricted globally to high income countries: so why are evidenced based prevention and treatment programs for mental disorders so rare? Asian J Psychiatr 2014 Aug;10:71-74. [doi: 10.1016/j.ajp.2014.06.007] [Medline: 25042956]

12. Schröder J, Berger T, Westermann S, Klein JP, Moritz S. Internet interventions for depression: new developments. Dialogues Clin Neurosci 2016 Jun;18(2):203-212. [Medline: 27489460]

13. Ehrmann J, Lewis C, Lewis P. Homo ludens revisited. Yale French Stud 1968(41):31. [doi: 10.2307/2929664]

14. What can we learn from playing interactive games. In: Playing Video Games Motives, Responses, and Consequences. New York: Routledge; 2006.

15. van Berkel N, Goncalves J, Hosio S, Kostakos V. Gamification of mobile experience sampling improves data quality and quantity. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Sep 11;1(3):1-21. [doi: 10.1145/3130972]

16. Brown S, Vaughan C. Play: How it Shapes the Brain, Opens the Imagination, and Invigorates the Soul. New York: Avery; 2010.

17. McGonigal J. Reality is Broken Why Games Make Us Better and How They Can Change the World. New York: Random House; 2011.

18. Visch VT, Vegt NJ, Anderiesen H, van der Kooij K. Persuasive game design: a model and its definitions. In: Proceedings of the CHI2013 Workshop on Designing Gamification: Creating Gameful and Playful Experiences. 2013 Presented at: CHI2013 Workshop on Designing Gamification: Creating Gameful and Playful Experiences; Apr 27-May 2, 2013; Paris, France.

19. Keijsers L, Van Roekel E. Longitudinal methods in adolescent psychology. In: Reframing Adolescent Research. Oxfordshire, England, UK: Routledge; 2018.

20. Proudfoot J, Parker G, Hadzi Pavlovic D, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010 Dec 19;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]

21. Gindidis S, Stewart S, Roodenburg J. A systematic scoping review of adolescent mental health treatment using mobile apps. Adv Ment Health 2018 Sep 28;17(2):161-177. [doi: 10.1080/18387357.2018.1523680]

22. van Roekel E, Keijsers L, Chung JM. A review of current ambulatory assessment studies in adolescent samples and practical recommendations. J Res Adolesc 2019 Sep 17;29(3):560-577 [FREE Full text] [doi: 10.1111/jora.12471] [Medline: 31573762]

23. Larson R, Csikszentmihalyi M. The experience sampling method. In: Flow and the Foundations of Positive Psychology. Dordrecht: Springer; 2014.

24. Wichers M, Smit AC, Snippe E. Early warning signals based on momentary affect dynamics can expose nearby transitions in depression: a confirmatory single-subject time-series study. J Pers Oriented Res 2020;6(1):1-15 [FREE Full text] [doi: 10.17505/jpor.2020.22042] [Medline: 33569148]

25. Dietvorst E, Hiemstra M, Maciejewski D, van Roekel E, Bogt TF, Hillegers MH, et al. Grumpy or depressed? Disentangling typically developing adolescent mood from prodromal depression using experience sampling methods. J Adolesc 2021 Apr;88:25-35 [FREE Full text] [doi: 10.1016/j.adolescence.2021.01.009] [Medline: 33607507]

26. van Os J, Verhagen S, Marsman A, Peeters F, Bak M, Marcelis M, ESM-MERGE Investigators PhD, et al. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. Depress Anxiety 2017 Jun;34(6):481-493. [doi: 10.1002/da.22647] [Medline: 28544391]

27. Widdershoven RL, Wichers M, Kuppens P, Hartmann JA, Menne-Lothmann C, Simons CJ, et al. Effect of self-monitoring through experience sampling on emotion differentiation in depression. J Affect Disord 2019 Feb 01;244:71-77. [doi: 10.1016/j.jad.2018.10.092] [Medline: 30321767]

28. Havinga PJ, Maciejewski DF, Hartman CA, Hillegers MH, Schoevers RA, Penninx BW. Prevention programmes for children of parents with a mood/anxiety disorder: systematic review of existing programmes and meta-analysis of their efficacy. Br J Clin Psychol 2021 Jun;60(2):212-251 [FREE Full text] [doi: 10.1111/bjc.12277] [Medline: 33410149]

29. National Collaborating Centre for Mental Health. Depression: The NICE Guideline on the Treatment and Management of Depression in Adults. London, England: The British Psychological Society & The Royal College of Psychiatrists; 2010.

30. Djaouti D, Alvarez J, Jessel JP. Classifying serious games: the G/P/S model. In: Handbook of Research on Improving Learning and Motivation Through Educational Games: Multidisciplinary Approaches. Pennsylvania, USA: IGI Global; 2011.

31. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull 1985 Sep;98(2):310-357. [Medline: 3901065]

32. Cohen S. Social relationships and health. Am Psychol 2004 Nov;59(8):676-684. [doi: 10.1037/0003-066x.59.8.676]

33. Compas BE, Connor-Smith J, Jaser SS. Temperament, stress reactivity, and coping: implications for depression in childhood and adolescence. J Clin Child Adolescent Psychol 2004 Feb;33(1):21-31. [doi: 10.1207/s15374424jccp3301_3]

34. Fu Z, Burger H, Arjadi R, Nauta MH, Bockting CL. Explaining the efficacy of an internet-based behavioral activation intervention for major depression: a mechanistic study of a randomized-controlled trial. Clin Psychol Eur 2021 Sep 30;3(3):1-24 [FREE Full text] [doi: 10.32872/cpe.5467]

35. Eberle SG. The elements of play: toward a philosophy and a definition of play. Am J Play 2014;6(2):214-233 https://www.journalofplay.org/sites/www.journalofplay.org/files/pdf-articles/6-2-article-elements-of-play.pdf.

36. Scholten H, Granic I. Use of the principles of design thinking to address limitations of digital mental health interventions for youth: viewpoint. J Med Internet Res 2019 Jan 14;21(1):e11528 [FREE Full text] [doi: 10.2196/11528] [Medline: 31344671]

37. Cohen D, Lindvall M, Costa P. An introduction to agile methods. Adv Comput 2004;62:1-66. [doi: 10.1016/s0065-2458(03)62001-2]

38. Grow it! Corona project. OSF HOME. URL: https://osf.io/2at58/?view_only=b691104ecc3d45ad8b48e1bd60ad7125 [accessed 2022-02-01]

39. IBM Corp. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp; 2017.

40. Radomski AD, Wozney L, McGrath P, Huguet A, Hartling L, Dyson MP, et al. Design and delivery features that may improve the use of internet-based cognitive behavioral therapy for children and adolescents with anxiety: a realist literature synthesis with a persuasive systems design perspective. J Med Internet Res 2019 Feb 05;21(2):e11128 [FREE Full text] [doi: 10.2196/11128] [Medline: 30720436]

41. Ryan AM. The peer group as a context for the development of young adolescent motivation and achievement. Child Dev 2001;72(4):1135-1150. [doi: 10.1111/1467-8624.00338] [Medline: 11480938]

42. van Hoorn J, van Dijk E, Meuwese R, Rieffe C, Crone EA. Peer influence on prosocial behavior in adolescence. J Res Adolesc 2014 Sep 27;26(1):90-100. [doi: 10.1111/jora.12173]

43. Telzer EH, van Hoorn J, Rogers CR, Do KT. Social influence on positive youth development: a developmental neuroscience perspective. Adv Child Dev Behav 2018;54:215-258 [FREE Full text] [doi: 10.1016/bs.acdb.2017.10.003] [Medline: 29455864]

XSL•FO
RenderX

44. Legerstee JS, Garnefski N, Jellesma FC, Verhulst FC, Utens EM. Cognitive coping and childhood anxiety disorders. Eur Child Adolesc Psychiatry 2010 Feb;19(2):143-150. [doi: 10.1007/s00787-009-0051-6] [Medline: 19727903]

45. Connor-Smith JK, Compas BE, Wadsworth ME, Thomsen AH, Saltzman H. Responses to stress in adolescence: measurement of coping and involuntary stress responses. J Consult Clin Psychol 2000 Dec;68(6):976-992. [Medline: 11142550]

46. Grow it!. Google Play. URL: https://play.google.com/store/apps/details?id=nl.erasmusmc.growit [accessed 2022-02-24]

47. Grow it!. App Store. URL: https://apps.apple.com/nl/app/grow-it/id1496767455 [accessed 2022-02-24]

48. Bergin AD, Vallejos EP, Davies EB, Daley D, Ford T, Harold G, et al. Preventive digital mental health interventions for children and young people: a review of the design and reporting of research. NPJ Digit Med 2020;3:133 [FREE Full text] [doi: 10.1038/s41746-020-00339-7] [Medline: 33083568]

49. Marcell AV, Klein JD, Fischer I, Allan MJ, Kokotailo PK. Male adolescent use of health care services: where are the boys? J Adolescent Health 2002 Jan;30(1):35-43. [doi: 10.1016/s1054-139x(01)00319-6]

50. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]

51. Schreuder MJ, Hartman CA, George SV, Menne-Lothmann C, Decoster J, van Winkel R, et al. Early warning signals in psychopathology: what do they tell? BMC Med 2020 Oct 14;18(1):269 [FREE Full text] [doi: 10.1186/s12916-020-01742-3] [Medline: 33050891]

52. Thabrew H, Fleming T, Hetrick S, Merry S. Co-design of eHealth interventions with children and young people. Front Psychiatry 2018 Oct 18;9:481 [FREE Full text] [doi: 10.3389/fpsyt.2018.00481] [Medline: 30405450]

53. Riese H, von Klipstein L, Schoevers RA, van der Veen DC, Servaas MN. Personalized ESM monitoring and feedback to support psychological treatment for depression: a pragmatic randomized controlled trial (Therap-i). BMC Psychiatry 2021 Mar 10;21(1):143 [FREE Full text] [doi: 10.1186/s12888-021-03123-3] [Medline: 33691647]

54. Dietvorst E, Legerstee JS, Vreeker A, Koval S, Mens M, Keijsers L, et al. The Grow It! Serious game: longitudinal changes in adolescent well-being during COVID-19. Open Science Framework 2021 preregistered(forthcoming) [FREE Full text]

55. Mens M, Keijsers L, Dietvorst E, Koval S, Legerstee JS, Hillegers MH. Mapping Mood during the COVID-10 pandemic in Dutch adolescents and youngsters. Open Science Framework 2021 preregistered (forthcoming) [FREE Full text]

Abbreviations

CBT: cognitive behavioral therapy
ESM: experience sampling method
mHealth: mobile health
MVP: minimal viable product

Original Paper

Feasibility and Preliminary Efficacy of Web-Based and Mobile Interventions for Common Mental Health Problems in Working Adults: Multi-Arm Randomized Pilot Trial

Marcos Economides[1], BSc, MPhil, PhD; Heather Bolton[1], MA, MSc, DClinPsy; Rhian Male[1], BSc, MSc; Kate Cavanagh[2], BA, DPhil, DClinPsy

[1]Unmind Ltd, London, United Kingdom

[2]School of Psychology, University of Sussex, East Sussex, United Kingdom

Corresponding Author:
Marcos Economides, BSc, MPhil, PhD
Unmind Ltd
180 Borough High St
London, SE1 1LB
United Kingdom
Phone: 44 2078664050
Email: marcos.economides@unmind.com

Abstract

Background: There is growing interest in digital platforms as a means of implementing scalable, accessible, and cost-effective mental health interventions in the workplace. However, little is known about the efficacy of such interventions when delivered to employee groups.

Objective: This study aims to evaluate the feasibility and preliminary efficacy of a digital mental health platform for the workplace, which incorporates evidence-based practices such as cognitive behavioral therapy and acceptance and commitment therapy. A total of 3 brief, unguided interventions designed to address stress, anxiety, and resilience, respectively, are evaluated. The primary aim is to determine the feasibility of the study methods and interventions in preparation for a definitive randomized controlled trial.

Methods: The study used a fully remote, parallel, multi-arm, external pilot randomized controlled trial, with 3 intervention arms and a no-intervention control group. Participants were working adults representative of the general UK population with respect to age, sex, and ethnicity who were recruited from a web-based participant platform. Primary outcomes included objective and self-report measures of feasibility, acceptability, engagement, transferability, relevance, and negative effects. Secondary outcomes included 4 self-report measures of mental health and well-being, completed at baseline (time point 0 [t0]), postintervention (time point 1 [t1]), and the 1-month follow-up (time point 2 [t2]). Secondary outcomes were analyzed via linear mixed-effects models using intention-to-treat principles. Preregistered criteria for progression to a definitive trial were evaluated.

Results: Data were collected between January and March of 2021. A total of 383 working adult participants meeting trial eligibility were randomized, of whom 356 (93%) were retained at t2. Objective engagement data showed that 67.8% (196/289) of participants randomized to an intervention arm completed their intervention. Overall, 87.1% (203/233) of participants reported being *satisfied* or *very satisfied* with their intervention and rated the quality of their intervention as *good* or *excellent*. All intervention groups reported significantly greater improvements than the control group on at least one secondary outcome at t1, with between-group Hedges g effect sizes for the pooled interventions ranging from 0.25 (95% CI 0.05-0.46) to 0.43 (95% CI 0.23-0.64). All the improvements were maintained at t2.

Conclusions: The study methods were feasible, and all preregistered criteria for progression to a definitive trial were met. Several minor protocol amendments were noted. Preliminary efficacy findings suggest that the study interventions may result in improved mental health outcomes when offered to working adults.

Trial Registration: ISRCTN Registry 80309011; http://www.isrctn.com/ISRCTN80309011

XSL•FO
RenderX

Introduction

Background and Rationale

Mental illness affects hundreds of millions of people worldwide, resulting in decreased quality of life, family and community disruption, increased health care costs, and a significant economic burden for employers [1,2]. Employee performance, rates of illness, absenteeism, and staff turnover are all affected by employees' mental health status. In the United Kingdom, workplace mental health problems result in an estimated 70 million lost workdays and a total cost of up to £45 billion (US $61 billion) each year for businesses [3]. This is compounded by an estimated global treatment gap of >50% for people with mental health disorders [4,5].

There is growing interest in web and smartphone apps as a means of increasing the reach of mental health and well-being interventions [6,7]. Digital platforms can offer a broad range of content within a standardized environment that is interactive and dynamic while also being widely accessible, cost-efficient, and nonstigmatizing. With fewer access barriers, digital platforms also have the potential to offer a preventative solution to common mental health problems by facilitating sustained, proactive engagement [8,9]. Such platforms can vary widely in their means of delivery (web vs mobile app), the core therapeutic approach they use (with cognitive behavioral therapy [CBT], mindfulness meditation [MM], and positive psychology being common), and the duration and format of their content.

There is now convincing evidence for the effectiveness of digital interventions when delivered in health and community settings [10], as well as emerging evidence that they may be effective when delivered in occupational settings [11,12]. Previous meta-analyses have found small positive effects on psychological well-being (Hedges g=0.37) and work effectiveness (Hedges g=0.25) [11] and small to moderate effects on common mental health outcomes, such as stress (Hedges g=0.54), anxiety (Hedges g=0.34), and symptoms of depression (Hedges g=0.30) [12]. However, the current evidence base is limited by considerable heterogeneity across studies and an insufficient number of high-quality trials. Moreover, only a fraction of for-profit mental health apps (MHapps) are supported by empirical evidence [13], with added concerns that such platforms are frequently characterized by low adherence [14-16]. Together, these suggest the need for further research.

In this study, we conduct an external pilot randomized controlled trial (RCT) as part of the initial testing of *Unmind*—a novel digital mental health platform for the workplace. *Unmind* provides employees with tools to help them track, maintain, and improve their mental health and well-being. It features a broad range of content that draws on multiple evidence-based approaches such as CBT [17], MM [18], behavioral activation [19], acceptance and commitment therapy (ACT [20]), and positive psychology [21]. Central to the platform are individual learning and development courses (known as *Series*) designed to address specific topics of mental health and well-being. *Series*

are short, standalone interventions, typically ranging between 5 and 7 sessions, each of approximately 10 minutes in duration, and can feature a mix of audio and video content, infographics, and interactions with a chatbot.

Study Objective

Consistent with recent guidelines on pilot trials [22,23], the primary aim of this study is to evaluate the feasibility of the study methods, and 3 separate *Unmind Series* that address the topics of stress, anxiety, and resilience, respectively, in preparation for a future definitive RCT. We chose to evaluate content relating to stress and anxiety as these are highly prevalent in the workplace [3] and have been extensively studied in previous evaluations of MHapps [24-26], allowing for a comparison of the current findings to previous evidence. In addition, we chose to evaluate content relating to resilience, as evidence suggests that it plays an important role in the prevention of mental health problems [27] and thus may be integral to the effectiveness of a preventative platform. A secondary aim is to establish the preliminary efficacy of each intervention with respect to self-report measures of stress, anxiety, symptoms of depression, and resilience, including establishing between-group effect sizes and 95% CIs (for each intervention compared with the control group). Although depression was not a specific target of any of the study interventions, we chose to include it as an outcome as it is highly comorbid with stress and anxiety [28,29] and a common problem in workplace settings [3]. Finally, we also aim to report on the combined effects of all interventions compared with the control group.

As the *Unmind* app comprises an extensive library of standalone interventions, it is important that each component of the app is evaluated. We chose to include 3 intervention arms in this study as this is more efficient than performing sequential 2-arm trials and increases the proportion of participants randomized to an intervention arm [30]. In addition, if a definitive RCT is warranted, an aim might be to evaluate whether each intervention arm has a greater effect on the specific outcome targeted by that intervention relative to the other intervention arms. Thus, this study uses a parallel, multi-arm, external pilot RCT design and recruited UK-based, community-dwelling, working adult participants who are randomly allocated to 1 of 3 intervention arms or to a no-intervention control group. We chose to implement a no-intervention control as (1) participants were not selected on the basis of poor mental health or seeking help for a problem, (2) recent evidence suggests that wait-list groups may introduce nocebo effects in psychotherapy trials [31], and (3) participants received monetary compensation for taking part.

The feasibility of each intervention arm is assessed via objective and self-reported outcomes capturing recruitment, retention, intervention uptake and adherence, acceptability, transferability, relevance, and negative effects. The preliminary efficacy of each intervention arm is assessed via self-report outcome measures delivered before (time point 0 [t0]) and after the

interventions (time point 1 [t1]; 2 weeks after t0) and at the 1-month follow-up (time point 2 [t2]). The results of this study are intended to inform whether a definitive RCT to evaluate the efficacy of each intervention arm is warranted and provide estimates of the parameters required for its design and implementation.

Methods

The authors followed the CONSORT (Consolidated Standards of Reporting Trials) 2010 guidelines [32] when preparing this study, including recent extensions to pilot trials [22] and multi-arm trials [33].

Trial Design

This study was a parallel, multi-arm, external pilot RCT with pre- (t0) and postintervention (t1; 2 weeks after t0) assessments and a 1-month follow-up (t2). Participants were randomly allocated to 1 of 3 brief, self-guided psychological interventions (*Series*) featured on the *Unmind* platform or to a no-intervention control group in a 1:1:1:1 allocation ratio. Participants were working adults recruited from the Prolific web-based recruitment platform [34], and the entire study was conducted on the web between January and March 2021. Of note, the study commenced several weeks after the start of a third national UK lockdown (in response to the SARS-CoV-2 pandemic), and t2 data were collected after the commencement of a phased easing of lockdown restrictions. The trial was preregistered at ISRCTN 80309011, and a full study protocol was preregistered at Open Science Framework in December 2020.

Ethics Approval

The trial received ethical approval from the University of Sussex sciences and technology research ethics committee (ER/KC226/2).

Participants

Participants were recruited via the Prolific web-based recruitment platform, which has been empirically tested across attributes such as participant response rates and data quality [35]. Inclusion criteria were (1) aged at least 18 years, (2) currently residing in the United Kingdom, (3) self-identifying as being in full- or part-time employment, (4) having an active account on Prolific, (5) having access to an internet connection via a smartphone or desktop device, and (6) being fluent in English. Prolific indicated that there were 50,978 eligible individuals at the time of conducting the study.

Prolific implements a prescreening system that allows researchers to screen for eligibility without implementing a screening questionnaire. Prolific also supports the recruitment of study samples representative of the national UK population with respect to age, sex, and ethnicity based on guidelines from the UK Office of National Statistics. This study drew upon this feature to maximize the generalizability of the findings.

Procedures

All study assessments were hosted on the Qualtrics survey platform [36], and participants were required to provide informed consent via a form built into each study assessment alongside a digital information sheet. All participants who completed the t0 assessment were invited to complete the t1 and t2 assessments via the Prolific recruitment platform. Participants were offered a £7 (US $9.50) incentive for completing each of the 3 study assessments (baseline, postintervention, and 1-month follow-up), delivered via Prolific. Participants randomized to one of the intervention arms received reminder messages on days 5 and 10 of the intervention (delivered via Prolific's anonymous inbox system), encouraging them to complete all intervention sessions. However, participant reimbursement was not contingent on intervention adherence.

Randomization

Randomization occurred at the end of t0 and was implemented via the Qualtrics *randomizer* feature, which uses block randomization to ensure balanced groups. It was not possible to blind the participants to group assignments. After randomization, participants assigned to one of the intervention arms were sent a message via Prolific's anonymous inbox system with instructions on how to access their intervention, including using a unique voucher code to sign up to the *Unmind* platform. The research team remained blind to group assignment for the duration of data collection but was unblinded during data analysis.

Interventions

Overview

Unmind is a digital platform designed to be used by working adults to measure, manage, and improve their mental health and well-being. It can be accessed via the web, mobile, or tablet (Android or iOS), and the *Unmind* smartphone app can be downloaded via the Apple or Google Play stores. The platform features a wide range of resources and content created by academics and clinicians with expertise in adult mental health, which are rooted in evidence-based practices such as CBT [17], MM [18], behavioral activation [19], ACT [20], and positive psychology [21].

Although *Unmind* includes a wide range of content and features, this study focused on evaluating *Series*. *Series* are brief, unguided learning and development courses, typically comprising between 5 and 7 sessions, each of approximately 10 minutes in duration, that are designed to be completed sequentially, and include a mix of audio and video content, infographics, and interaction with a chatbot (see Figure 1 for example screenshots). Each *Series* focuses on a specific symptom, topic, trait, or behavior related to mental health and typically uses a key therapeutic approach, such as CBT, MM, or ACT. *Series* are designed to provide both reactive support (to manage or address an existing problem) and proactive support (to prevent the onset of a future mental health problem).

Figure 1. Screenshots of the *Unmind* smartphone app showing the *Series* tab (panel A) and examples from the *combatting stress* intervention (panels B-D).
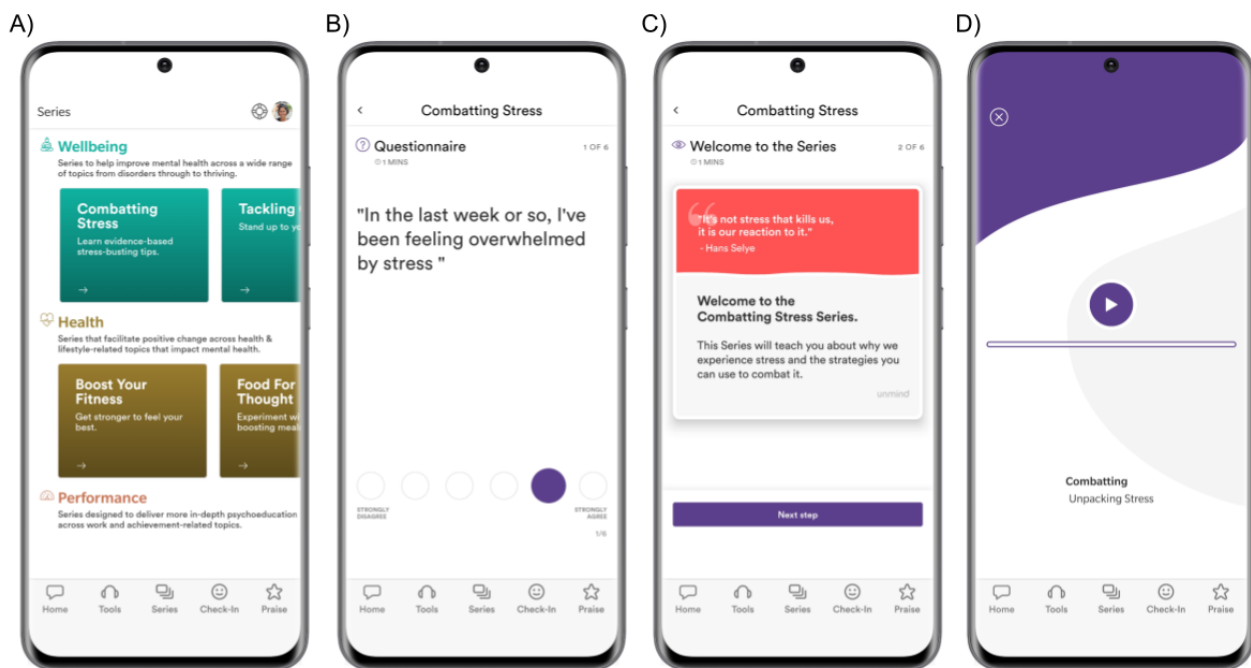
The intervention arms in this study comprised 3 individual *Series* designed to address stress, anxiety, and resilience, respectively. For the purposes of the study, participants were instructed to only engage with their allocated intervention, despite having access to the full *Unmind* platform, and were excluded from standard email campaigns that encourage interaction with content not evaluated in this study. Participants had 2 weeks to complete their allocated intervention and were free to progress through the intervention at their own pace. A description of each intervention arm is provided in the following sections.

Combatting Stress

This intervention draws upon CBT and ACT techniques and is designed to help users better manage their day-to-day stressors. Over the course of 7 sessions, it provides psychoeducation on stress and its physical manifestations, helps users spot personal triggers, and explores different approaches to coping. It also introduces the idea of acceptance. Users are taught stress management techniques and are encouraged to practice between sessions.

Working With Worry

This intervention is underpinned by theoretical models of generalized anxiety disorder (GAD), although it is targeted at users who identify as worriers rather than those meeting any predefined criteria for a diagnosis of GAD. Content spans 7 sessions and covers key elements of CBT, including tolerance of uncertainty, challenging worry beliefs, problem solving, and working with imagery. It also encourages users to apply evidence-based techniques, including relaxation and attentional focus.

Building Resilience

This intervention aims to help users apply evidence-based techniques to aid the cultivation of essential qualities of personal resilience, drawing upon CBT and ACT. Over the course of 7 sessions, learning covers topics such as honing strengths, facing challenges, and tolerating discomfort. It also explores aspects such as coping styles and realistic optimism. The *Series* encourages users to increase their self-awareness and guides them to build a personal resilience plan.

In each *Series*, learning is optimized by the use of a chatbot to allow note-taking and aid reflection, as well as the use of short recap videos at the beginning of each new session. Each *Series* also has its own accompanying brief handbook, which is emailed to participants as a PDF file on starting their first session. Handbooks contain a summary of key learning points and infographics from the *Series* and include space for participants to write any further reflections on their learning from each session.

For the purposes of this study, participants were instructed not to engage with other content and features included in the *Unmind* app (and not described here) so that the feasibility of the study interventions could be assessed in isolation. The app also includes a *Help* page containing information and resources for key well-being topics and signposting to urgent problems. Participants had access to versions 2.56.0 to 2.59.0 of the *Unmind* app, and no major app changes or updates were launched during the 2-week study period.

Outcomes

Overview

Participant demographics and other variables were captured at t0, including whether each participant had engaged with therapy or counseling, and a mental health or well-being app within 6

months before taking part in the study. Participants were also asked to rate the extent to which they agreed with the statement, "Do you agree that it's important for people to look after their mental health and wellbeing?" on a 5-point Likert scale from *strongly agree* to *strongly disagree*.

Primary Outcome Measures

Recent guidelines suggest that complex health interventions should be feasible, acceptable, engaging, transferable to other settings, and relevant [37]. In addition, psychological interventions should be evaluated for negative effects [38]. Therefore, preregistered primary outcomes included the following:

1. Feasibility: recruitment, intervention uptake, and retention (at t1 and t2)
2. Acceptability: intervention adherence and completion rates, participant satisfaction, and reasons for discontinuing the intervention
3. Engagement: average sessions completed and 3 questions adapted from the Mobile App Rating Scale [39]
4. Transferability: 1 question adapted from the Mobile App Rating Scale
5. Relevance: 1 question assessing subjective relevance
6. Negative effects: 1 question adapted from recent guidelines on assessing negative effects [40] and the proportion of participants that reliably deteriorated across all secondary outcome measures.

Outcomes were measured through a combination of objective data (captured by the *Unmind* platform) and self-reported data captured at t1 (Multimedia Appendix 1).

Secondary Outcome Measures

Preregistered secondary outcomes included self-report measures capturing symptoms of common mental health problems.

The Perceived Stress Scale-10

The Perceived Stress Scale (PSS) is a 10-item scale that asks respondents to rate how often they feel or think that their lives are unpredictable, uncontrollable, and overloaded on a 5-point Likert scale from 0 (never) to 4 (very often) [41]. Total scores range from 0 to 40, with higher scores indicating greater perceived stress. The PSS has a Cronbach α >.70 across 12 individual studies (and .91 in this study at t0) and good test–retest reliability across 4 individual studies [42]. The original scale uses a 1-month reporting period; however, this has been shortened in several previous studies [43,44], and this study used a 2-week reporting period.

The GAD-7 Scale

The GAD-7 is a 7-item scale used to screen for the presence and severity of an anxiety disorder [45]. Participants rate each item on a 4-point Likert scale from 0 (not at all) to 3 (nearly every day), with total scores ranging from 0 to 21. A score ≥10 is suggestive of the presence of anxiety, and scores of 5, 10, and 15 are taken as cutoff points for mild, moderate, and severe anxiety, respectively. The GAD-7 has excellent reliability and internal consistency (Cronbach α of .89 in the original validation and .91 in this study at t0) and has been validated in both the general population and primary care settings [45,46].

The Patient Health Questionnaire-8

The Patient Health Questionnaire (PHQ)-8 is an 8-item scale derived from the PHQ-9, which screens for the presence and severity of depression [47]. The PHQ-8 omits an item that assesses suicidal ideation and is more appropriate for use in nonclinical samples and settings [48]. The response options are equivalent to those of the GAD-7, with total scores ranging from 0 to 24. A score ≥10 suggests the presence of depression, and scores of 5, 10, 15, and 20 are taken as cutoff points for mild, moderate, moderately severe, and severe depression, respectively. The PHQ-8 has excellent internal consistency (Cronbach α of .89 in primary care settings and .88 in this study at t0) and excellent test–retest reliability [49].

Brief Resilience Scale

The Brief Resilience Scale (BRS) is a short, 6-item scale designed to assess people's ability to bounce back or recover after stressful events [50]. Participants rate each item on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree) or the reverse for negatively worded items. The BRS is scored by reverse coding items 2, 4, and 6 and computing the mean of the 6 items. The creators of the scale have suggested that scores <3 be interpreted as *low* resilience and scores ≥4.3 be interpreted as *high* resilience [51]. The BRS displays good internal consistency (Cronbach α=.80-.91 and .91 in this study at t0) and test–retest reliability.

Progression Criteria

As per formal guidelines [52], preregistered progression criteria were defined as follows: (1) full study recruitment within 1 month; (2) at least 30% intervention completion rates based on a previous meta-analysis of adherence to unguided psychological interventions [53]; (3) at least 75% adherence to protocol instructions (defined as the proportion of participants who refrain from engaging with ≥1 *Series* session outside of their allocated *Series*); (4) at least 50% of participants reporting being *satisfied* or *very satisfied* with the intervention and rating the quality of the intervention as *good* or *excellent*; and (5) the 95% CI on between-group effect sizes for secondary outcomes including at least a small effect (Hedges g=0.2) for ≥1 outcome measures.

Progression criteria were considered for each intervention arm individually.

Sample Size

This study was powered for CIs on the feasibility outcomes. A sample size calculation indicated that approximately 100 participants were required to estimate feasibility outcomes with a margin of error ≤10% (based on a conservative population proportion of 50% for retention and adherence, and a 95% CI). This is consistent with previous guidelines suggesting that 60 to 100 participants per intervention arm are optimal for estimating binary outcomes in pilot RCTs [54]. Therefore, we aimed to recruit 400 participants in total.

Statistical Methods

The results from all preregistered primary and secondary measures are reported. Minor deviations from the preregistered

data analysis plan are reported in Multimedia Appendix 2 [40,41,55].

Primary Analyses

Descriptive statistics were used to report primary outcomes. Categorical data were reported as proportions in each response category, and Fisher exact test of independence was used to compare responses between intervention arms (with *P* values computed using Monte Carlo simulation and 2000 iterations). Where tests were significant, post hoc pairwise comparisons between study arms were performed (using false discovery rate methods to adjust *P* values).

Objective in-app usage data were provided by *Unmind*. For simplicity, intervention sessions were only characterized as *complete* if all components of the session were played. Descriptive statistics were used to characterize engagement and stratify participants according to whether they completed, started but did not complete, or failed to start their allocated intervention.

We computed the proportion of participants who self-reported reliable deterioration in mental health scores from t0 to t1, and t1 to t2, based on an estimate of the reliable change index for each outcome measure. The reliable change index was computed based on methods provided by Jacobson and Truax [56], using Cronbach α as a measure of reliability and an α level of .05 (Multimedia Appendix 3). Participants with missing data or those who were unable to reliably deteriorate based on t0 scores were excluded from this analysis.

Secondary Analyses

Secondary outcome measures were analyzed using both intention-to-treat (ITT) and per-protocol (PP) approaches. For the ITT analysis, all participants with complete t0 data were included, regardless of intervention adherence and any deviation from instructions. Participants were excluded from the PP analysis if they failed to complete all 7 intervention sessions, if they started an *Unmind Series* outside of their allocated intervention, or if they were lost to follow-up at t1. As findings from the PP analysis were largely equivalent to ITT, we opted to omit these results (although a comparison of effect sizes is reported in Multimedia Appendix 4).

Analyses were performed using linear mixed-effects models (LMMs) with restricted information maximum likelihood estimation (via the *lme4* package in R [57]). Each model included a within-subject factor *time* (with levels: t0, t1, and t2), a between-subject factor *group* (*combatting stress*, *working with worry*, *building resilience*, or control), their interaction as fixed effects, and a separate baseline for each participant. Time was modeled as a categorical factor. Model residuals were checked via Q–Q plots to assess model assumptions and goodness of fit. For each outcome, we reported (1) the estimated marginal means (EMMs) with 95% CIs for each time point and intervention arm, (2) *P* values for within-group contrasts

comparing changes from t0 to t1 and t0 to t2, and (3) between-group contrasts (with 95% CIs) comparing changes from t0 to t1 and t0 to t2 for each intervention arm (and all intervention arms combined) relative to the control group (with both unadjusted and Tukey-adjusted *P* values). *P* values <.05 were considered significant. We also report a standardized effect size (Hedges *g* with 95% CI) for each between-group contrast. Hedges *g* was calculated using EMMs (as opposed to raw data, which require the use of complete cases only) and pooled SDs. The 95% CIs were calculated using equations 15 and 16 from Nakagawa and Cuthill [58]. *P* values were reported to a maximum of 3 decimal places, with values <.001 reported as *P*<.001.

Subgroup Analyses

Subgroup analyses were performed to examine changes in secondary outcome measures for participants who self-reported having at least mild symptoms at t0 or at least moderately low resilience. Thresholds for subgroup analyses were as follows: a score ≥16 on the PSS, ≥5 on the GAD-7, ≥5 on the PHQ-8, and <3 on the BRS. For simplicity, we report a comparison of Hedges *g* effect sizes for these subgroups versus the ITT analysis but omit the full output of each LMM. In addition, we analyzed the intervention feedback ratings for these subgroups separately. As the findings were similar to the ITT sample, these are reported in Multimedia Appendix 5.

Finally, multivariate logistic regression was conducted on the intervention group data only to explore whether any baseline variables were predictive of intervention completion (defined as *1* for randomized participants who completed all sessions of their allocated intervention and *0* for all other randomized participants). Predictor variables included all demographic variables and other baseline characteristics, as well as all self-report secondary outcome measures at t0. For categorical predictors, categories that included <10 observations were dropped from the regression analysis.

Results

Participants

Participant demographics and other baseline variables are presented in Table 1. The mean age of the participants was 44.6 (SD 14.3) years, 52% (199/383) were female, and 81.2% (311/383) were White, suggesting that the study sample was broadly representative of the general UK population [59]. All participants were employed (part-time: 94/383, 24.5%; self-employed: 49/383, 12.8%) across a broad range of industries and most had not used an MHapp (306/383, 79.9%) or not engaged in talking therapy (352/383, 91.9%) in a 6-month period before taking part in the study. Almost all participants (372/383, 97.1%) either *agreed* or *strongly agreed* with the statement, "It's important for people to look after their mental health and wellbeing."

Table 1. Participant demographics and baseline variables (N=383).

Variable	Overall	Study arm			
		Control (n=94)	CS[a] (n=94)	WW[b] (n=97)	BR[c] (n=98)
Age (years), mean (SD; range)	44.6 (14.3; 18-75)	45.6 (14.2; 18-69)	44.8 (14.3; 19-75)	43.6 (14.7; 19-72)	44.7 (14.3; 18-69)
Sex, n (%)					
Female	199 (52)	53 (56.4)	41 (43.6)	55 (56.7)	50 (51)
Male	184 (48)	41 (43.6)	53 (56.4)	42 (43.3)	48 (49)
Ethnicity, n (%)					
White[d]	311 (81.2)	76 (80.9)	79 (84)	74 (76.3)	82 (83.7)
Black[e]	19 (5)	4 (4.3)	2 (2.1)	7 (7.2)	6 (6.1)
Mixed or multiple	12 (3.1)	2 (2.1)	3 (3.2)	5 (5.2)	2 (2)
Asian[f]	35 (9.1)	11 (11.7)	9 (9.6)	9 (9.3)	6 (6.1)
Other[g]	6 (1.6)	1 (1.1)	1 (1.1)	2 (2.1)	2 (2)
Employment, n (%)					
Full-time	238 (62.1)	60 (63.8)	63 (67)	55 (56.7)	60 (61.2)
Part-time	94 (24.5)	26 (27.7)	18 (19.1)	28 (28.9)	22 (22.4)
Self-employed	49 (12.8)	8 (8.5)	13 (13.8)	13 (13.4)	15 (15.3)
Other	2 (0.5)	0 (0)	0 (0)	1 (1)	1 (1)
Industry, n (%)					
Agriculture, forestry, or mining	1 (0.3)	1 (1.1)	0 (0)	0 (0)	0 (0)
Industrials	22 (5.7)	4 (4.3)	9 (9.6)	3 (3.1)	6 (6.1)
Energy or utilities	7 (1.8)	2 (2.1)	1 (1.1)	2 (2.1)	2 (2)
Transport or logistics	17 (4.4)	2 (2.1)	5 (5.3)	6 (6.2)	4 (4.1)
Media or creative industries	24 (6.3)	6 (6.4)	9 (9.6)	4 (4.1)	5 (5.1)
Data or telecommunication	18 (4.7)	5 (5.3)	5 (5.3)	2 (2.1)	6 (6.1)
Health care	40 (10.4)	13 (13.8)	6 (6.4)	12 (12.4)	9 (9.2)
Education	63 (16.4)	17 (18.1)	17 (18.1)	12 (12.4)	17 (17.3)
Life sciences	4 (1)	0 (0)	1 (1.1)	2 (2.1)	1 (1)
Retail	32 (8.4)	2 (2.1)	11 (11.7)	8 (8.2)	11 (11.2)
Hospitality, leisure, or travel	22 (5.7)	8 (8.5)	7 (7.4)	3 (3.1)	4 (4.1)
Public or social service	30 (7.8)	8 (8.5)	6 (6.4)	7 (7.2)	9 (9.2)
Finances, insurance, or real estate	22 (5.7)	5 (5.3)	4 (4.3)	8 (8.2)	5 (5.1)
Professional services	25 (6.5)	7 (7.4)	4 (4.3)	8 (8.2)	6 (6.1)
Other	56 (14.6)	14 (14.9)	9 (9.6)	20 (20.6)	13 (13.3)
Education, n (%)					
None	4 (1)	0 (0)	2 (2.1)	1 (1)	1 (1)
High school	138 (36)	30 (31.9)	33 (35.1)	36 (37.1)	39 (39.8)
Undergraduate degree	172 (44.9)	46 (48.9)	41 (43.6)	42 (43.3)	43 (43.9)
Postgraduate degree	69 (18)	18 (19.1)	18 (19.1)	18 (18.6)	15 (15.3)
MHapp[h] use (6 months), n (%)					
Yes	74 (19.3)	15 (16.0)	17 (18.1)	23 (23.7)	19 (19.4)
No	306 (79.9)	78 (83.0)	76 (80.9)	73 (75.3)	79 (80.6)
Maybe	3 (0.8)	1 (1.1)	1 (1.1)	1 (1)	0 (0)

XSL•FO

RenderX

Variable	Overall	Study arm			
		Control (n=94)	CS[a] (n=94)	WW[b] (n=97)	BR[c] (n=98)
Therapy (6 months), n (%)					
Yes	31 (8.1)	6 (6.4)	3 (3.2)	12 (12.4)	10 (10.2)
No	352 (91.9)	88 (93.6)	91 (96.8)	85 (87.6)	88 (89.8)
Proactive MH[i] care important, n (%)					
Strongly disagree	6 (1.6)	0 (0)	2 (2.1)	1 (1)	3 (3.1)
Disagree	1 (0.3)	1 (1.1)	0 (0)	0 (0)	0 (0)
Neither	4 (1)	0 (0)	1 (1.1)	1 (1)	2 (2)
Agree	75 (19.6)	19 (20.2)	25 (26.6)	15 (15.5)	16 (16.3)
Strongly agree	297 (77.5)	74 (78.7)	66 (70.2)	80 (82.5)	77 (78.6)

[a]CS: combatting stress.

[b]WW: working with worry.

[c]BR: building resilience.

[d]White British and other British.

[e]African, Caribbean, and Black British.

[f]Chinese, Indian, Bangladeshi, Pakistani, and other Asian.

[g]Arabian or any other ethnicity.

[h]MHapp: mental health app.

[i]MH: mental health.

Patient-reported outcome scores suggested that participants were, on average, experiencing mild symptoms of depression and anxiety at t0 (mean PHQ-8 6.9, SD 5.2; mean GAD-7 6.5, SD 5.1). The proportion of participants scoring above the cutoff for mild symptoms was 59.3% (227/383) for anxiety (GAD-7≥5) and 59.8% (229/383) for depression (PHQ-8≥5), whereas the proportion scoring above the cutoff for moderate symptoms was 26.9% (101/383) for anxiety (GAD-7≥10) and 28.2% (108/383) for depression (PHQ-8≥10). Self-reported stress levels were approximately consistent with population norms (mean PSS 17.0, SD 7.7 [55]).

Primary Outcomes

Enrollment and Retention

The study was enrolled in January 2021 within 48 hours of launching the study advert. Figure 2 shows the participant flow through the trial. One of the participants withdrew consent after randomization, and 4% (16/400) of participants reported not being employed at t0 (in contrast to their prescreening responses) and were excluded from all analyses. Of the remaining 383 eligible participants, 367 (95.8%) completed an assessment at t1, and 356 (93%) completed an assessment at t2. Retention rates at t2 significantly differed across the intervention arms (*P*=.02; control 97.9%, *combatting stress* 93.6%, *working with worry* 93.8%, and *building resilience* 86.7%). Pairwise post hoc comparisons suggested significantly lower retention for *building resilience* than the control arm (adjusted *P*=.03); however, no other comparisons were significant. All groups exceeded the prespecified minimum retention for progression to a definitive trial.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow of participants through the study. t0: time point 0; t1: time point 1; t2: time point 2.
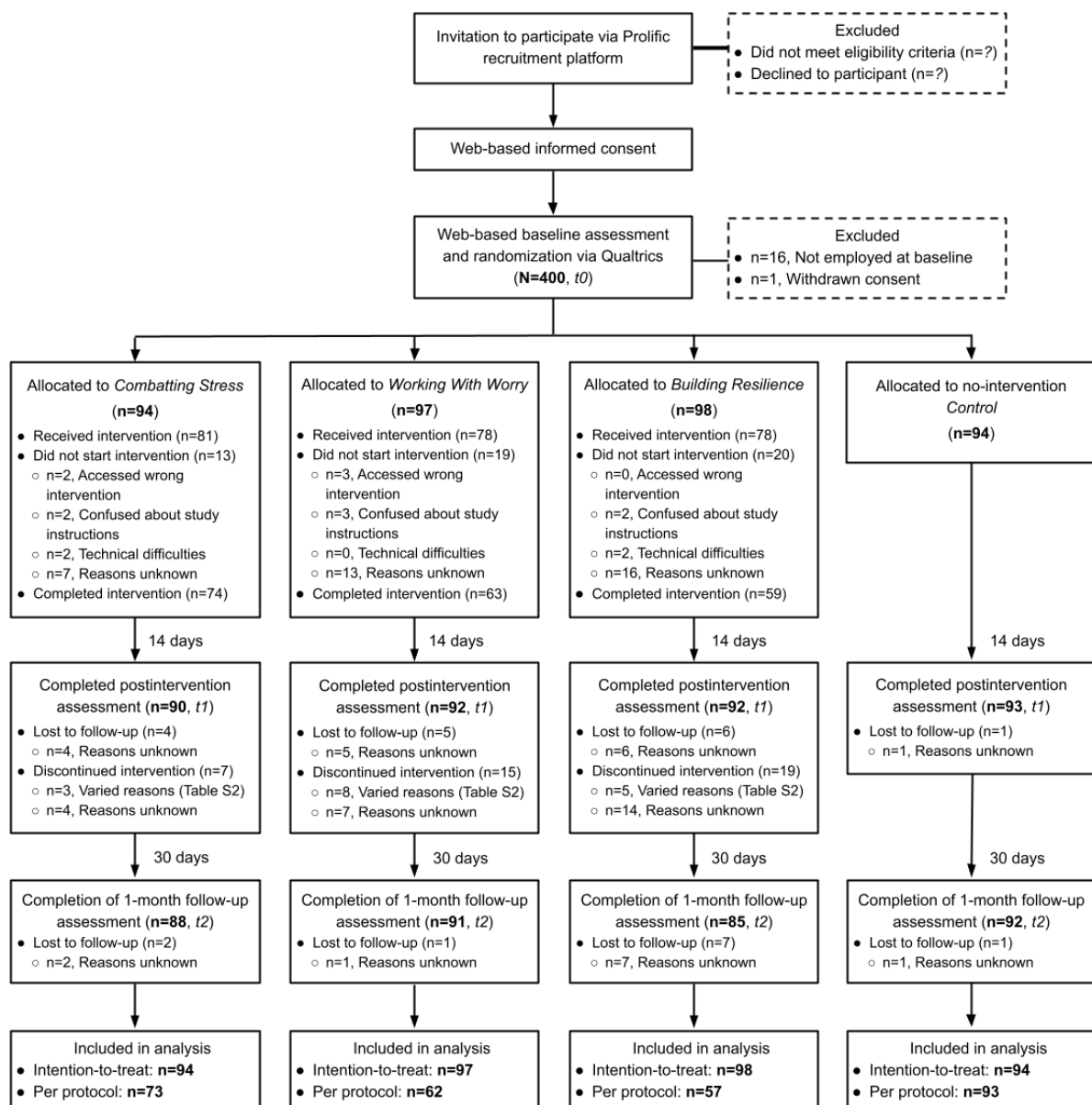
Engagement and Adherence

Overview

A summary of the intervention engagement is shown in Table 2. Of the 289 participants randomized to an intervention, 237 (82%) started their allocated intervention, and 196 (67.8%) completed all intervention sessions. Of those who completed at least one session, 82.7% (196/289) proceeded to complete all sessions, which differed across intervention arms (*P*=.02). Pairwise post hoc comparisons suggested significantly higher completion for *combatting stress* (74/81, 91.4%) than for *building resilience* (59/78, 75.6%; adjusted *P*=.03) but not *working with worry* (63/78, 80.8%). Of the 289 participants randomized to an intervention, 47 (16.3%) did not start their intervention, and 5 (1.7%) incorrectly engaged with an intervention outside of the one they were allocated. Participants took an average of 6.38 (SD 4.10, range 0-15) days to complete all 7 intervention sessions, which did not significantly differ

across intervention arms ($F_{2,193}$=1.58; *P*=.21). Those who started but did not complete their allocated intervention completed a median of 3 out of 7 sessions (mean 3.24, SD 1.96, range 1-6). A summary of participants' self-reported reasons for not starting or discontinuing their intervention can be found in Multimedia Appendix 6. The most common reasons for lack of engagement included insufficient time or technical difficulties (although these data were missing for participants who self-reported completing their intervention, regardless of objective adherence).

Overall, participants who objectively completed at least one intervention session self-reported spending a mean of 60.1 (median 60, SD 29.1, minimum 10, maximum 210) minutes on the *Unmind* platform, which differed significantly across groups ($F_{2,231}$=3.12; *P*=.046). Post hoc Tukey tests suggested that participants in the *combatting stress* arm reported spending more time on the platform than participants in the *building resilience* arm (mean difference 11.49, 95% CI 0.64-22.34;

adjusted *P*=.04) but not the *working with worry* arm (adjusted *P*=.40). Of the 274 participants who completed an assessment at t1, 177 (64.6%) reported receiving a handbook via email (*combatting stress* 58/90, 64%, *working with worry* 61/92, 66%, and *building resilience* 58/92, 63%) while completing their intervention. Of these 177 participants, 45 (25.4%) reported reading the entire handbook (*combatting stress* 12/58, 21%,

working with worry 16/61, 26%, and *building resilience* 17/58, 29%), and 84 (47.5%) reported reading some of the handbook (*combatting stress* 27/58, 47%; *working with worry* 31/61, 51%; and *building resilience* 26/58, 45%). Overall, 4.2% (12/289) of participants deviated from the study instructions by engaging with ≥1 session outside their allocated intervention arm.

Table 2. Intervention adherence and engagement across the 3 intervention arms (N=289).

Variable	Overall	Study arm		
		CS[a] (n=94)	WW[b] (n=97)	BR[c] (n=98)
Completers				
Value, n	196	74	63	59
Percentage of those randomized (95% CI)	67.8 (62.1-73.2)	78.7 (69.1-86.5)	64.9 (54.6-74.4)	60.2 (49.8-70)
Percentage of those starting intervention (95% CI)	82.7 (77.3-87.3)	91.4 (83.0-96.5)	80.8 (70.3-88.8)	75.6 (64.6-84.7)
Days taken to complete intervention, mean (SD; range)	6.38 (4.10; 0-15)	6.59 (3.94; 0-15)	6.86 (3.85; 0-14)	5.61 (4.51; 0-14)
Partial completers				
Value, n	41	7	15	19
Percentage of those randomized (95% CI)	14.2 (10.4-18.8)	7.45 (3.1-14.7)	15.5 (8.9-24.2)	19.4 (12.1-28.6)
Number of sessions completed by partial completers				
Values, mean (SD)	3.24 (1.96)	2.71 (1.80)	2.80 (1.47)	3.79 (2.27)
Values, median (range)	3 (1-6)	3 (1-5)	3 (1-5)	4 (1-6)
Did not start intervention				
Value, n	47	11	16	20
Percentage of those randomized (95% CI)	16.3 (12.2-21)	11.7 (6-20)	16.5 (9.7-25.4)	20.4 (12.9-29.7)
Engaged only with nonassigned intervention				
Value, n	5	2	3	0
Percentage of those randomized (95% CI)	1.7 (0.6-4)	2.1 (0.3-7.5)	3.1 (0.6-8.8)	0 (0-3.7)
Engaged with assigned and nonassigned interventions				
Value, n	12	4	5	3
Percentage of those randomized (95% CI)	4.2 (2.2-7.1)	4.3 (1.2-10.5)	5.2 (1.7-11.6)	3.1 (0.6-8.7)

[a]CS: combatting stress.

[b]WW: working with worry.

[c]BR: building resilience.

Satisfaction and Feedback Ratings

Table 3 shows a summary of feedback ratings from participants who were retained at t1 and who also completed at least one intervention session based on objective app use (233/383, 60.8%; see Multimedia Appendix 1 for feedback questions). Importantly, most (203/233, 87.1%) participants were either *satisfied* (99/233, 42.5%) or *very satisfied* (104/233, 44.6%) with their intervention and rated the quality of their intervention as either *good* (96/233, 41.2%) or *excellent* (107/233, 45.9%). Feedback did not significantly differ across the intervention arms, except for the intervention quality (*P*=.02). Post hoc

pairwise comparisons suggested that ratings differed between *working with worry* and *combatting stress* (adjusted *P*=.02) and differed marginally between *combatting stress* and *building resilience* (adjusted *P*=.07). Compared with *building resilience*, participants in the *combatting stress* arm were more likely to rate the intervention as *good* than *okay* or *poor*, and participants in the working with *worry* arm reported the highest ratio of *excellent* to *good ratings*.

Feedback ratings from participants scoring above predefined cutoffs for inclusion in subgroup analyses were largely equivalent to the overall sample and are included in Multimedia Appendix 5.

Table 3. Postintervention feedback ratings from participants who were retained at time point 1 (t1), both overall and for each intervention arm (N=233).

Feedback ratings	Overall n (%)	Study arm n (%)			P value[a]
		CS[b] (n=79)	WW[c] (n=76)	BR[d] (n=78)	
Design of intervention					.16
Dull, not fun	1 (0.4)	0 (0)	0 (0)	1 (1.3)	
Mostly boring	14 (6)	1 (1.3)	7 (9.2)	6 (7.7)	
OK, fun enough	55 (23.6)	17 (21.5)	15 (19.7)	23 (29.5)	
Moderately interesting and fun	103 (44.2)	41 (51.9)	32 (42.1)	30 (38.5)	
Highly interesting and fun	60 (25.8)	20 (25.3)	22 (28.9)	18 (23.1)	
Content of intervention					.46
Dull, not fun	3 (1.3)	0 (0)	1 (1.3)	2 (2.6)	
Mostly boring	11 (4.7)	2 (2.5)	5 (6.6)	4 (5.1)	
OK, fun enough	46 (19.7)	13 (16.5)	13 (17.1)	20 (25.6)	
Moderately interesting and fun	94 (40.3)	35 (44.3)	28 (36.8)	31 (39.7)	
Highly interesting and fun	79 (33.9)	29 (36.7)	29 (38.2)	21 (26.9)	
Relevance of intervention					.59
Strongly disagree	7 (3)	1 (1.3)	3 (3.9)	3 (3.8)	
Disagree	16 (6.9)	3 (3.8)	7 (9.2)	6 (7.7)	
Neither agree nor disagree	40 (17.2)	14 (17.7)	9 (11.8)	17 (21.8)	
Agree	97 (41.6)	36 (45.6)	33 (43.4)	28 (35.9)	
Strongly agree	73 (31.3)	25 (31.6)	24 (31.6)	24 (30.8)	
Satisfaction with intervention					.71
Very dissatisfied	3 (1.3)	0 (0)	1 (1.3)	2 (2.6)	
Dissatisfied	8 (3.4)	2 (2.5)	3 (3.9)	3 (3.8)	
Neither satisfied nor dissatisfied	19 (8.2)	4 (5.1)	7 (9.2)	8 (10.3)	
Satisfied	99 (42.5)	36 (45.6)	28 (36.8)	35 (44.9)	
Very satisfied	104 (44.6)	37 (46.8)	37 (48.7)	30 (38.5)	
Quality of intervention					.02
Poor	5 (2.1)	0 (0)	3 (3.9)	2 (2.6)	
Okay	25 (10.7)	4 (5.1)	9 (11.8)	12 (15.4)	
Good	96 (41.2)	42 (53.2)	23 (30.3)	31 (39.7)	
Excellent	107 (45.9)	33 (41.8)	41 (53.9)	33 (42.3)	
Likelihood of recommending intervention					.13
I would not recommend it to anyone	7 (3)	1 (1.3)	2 (2.6)	4 (5.1)	
There are very few people I would recommend it to	29 (12.4)	5 (6.3)	11 (14.5)	13 (16.7)	
There are several people whom I would recommend it to	65 (27.9)	28 (35.4)	17 (22.4)	20 (25.6)	
There are many people I would recommend it to	74 (31.8)	20 (25.3)	29 (38.2)	25 (32.1)	
Definitely—I would recommend it to everyone	58 (24.9)	25 (31.6)	17 (22.4)	16 (20.5)	
Ease of use of *Unmind* app and intervention					.96
No (limited instructions, confusing, and complicated)	1 (0.4)	0 (0)	0 (0)	1 (1.3)	
Useable after a lot of time and effort	1 (0.4)	0 (0)	1 (1.3)	0 (0)	
Usable after some time and effort	12 (5.2)	4 (5.1)	4 (5.3)	4 (5.1)	
Easy to learn how to use	75 (32.2)	26 (32.9)	22 (28.9)	27 (34.6)	

Feedback ratings	Overall n (%)	Study arm n (%)			P value[a]
		CS[b] (n=79)	WW[c] (n=76)	BR[d] (n=78)	
Able to use app immediately	144 (61.8)	49 (62)	49 (64.5)	46 (59)	
Negative effects during the intervention					.11
Yes	2 (0.9)	0 (0)	2 (2.6)	0 (0)	
No	231 (99.1)	79 (100)	74 (97.4)	78 (100)	

[a]Fisher exact tests comparing ratings across study arms.

[b]CS: combatting stress.

[c]WW: working with worry.

[d]BR: building resilience.

Negative Effects

Of the 76 participants in the *working with worry* arm, 2 (3%) reported experiencing negative effects while completing their intervention. Qualitative feedback suggested that in both instances, this referred to frustration with the intervention (not understanding the content or not finding it useful). Table 4 shows the proportion of participants whose self-reported outcome scores reliably deteriorated from t0 to t1 and t1 to t2 for each study arm and each secondary outcome measure. Across all outcomes, the proportion of participants who self-reported reliable deterioration ranged from 1.1% to 8.9% for t0 to t1 and 2.4% to 12.4% for t1 to t2 (during which participants no longer had access to any interventions). These rates were largely equivalent between the intervention arms and the control group and are consistent with previous estimates that 5% to 10% of participants are expected to deteriorate following in-person psychotherapy [60]. Thus, the interventions in this study did not appear to be associated with symptom deterioration or other unwanted negative effects.

Table 4. Number and percentage of participants per study arm (and overall) that reliably deteriorated from time point 0 (t0) to time point 1 (t1) and t1 to time point 2 (t2) based on reliable change index for each secondary outcome measure.

Outcome	Overall, n (%)	Study arm, n (%)			
		Control	CS[a]	WW[b]	BR[c]
PSS[d]					
t1 to t0 (n=365)	15 (4.1)	4 (4.3)	8 (8.9)	1 (1.1)	2 (2.2)
t2 to t1 (n=349)	35 (10)	9 (9.9)	6 (7.1)	11 (12.4)	9 (10.7)
GAD-7[e]					
t1 to t0 (n=353)	18 (5.1)	4 (4.4)	7 (8.0)	1 (1.1)	6 (6.9)
t2 to t1 (n=340)	27 (7.9)	9 (10.2)	3 (3.6)	7 (8.4)	8 (9.8)
PHQ-8[f]					
t1 to t0 (n=359)	17 (4.7)	7 (7.5)	6 (6.7)	1 (1.1)	3 (3.4)
t2 to t1 (n=344)	13 (3.8)	3 (3.4)	2 (2.4)	4 (4.5)	4 (4.8)
BRS[g]					
t1 to t0 (n=356)	14 (3.9)	4 (4.4)	3 (3.4)	4 (4.4)	3 (3.5)
t2 to t1 (n=347)	22 (6.3)	5 (5.4)	6 (7.1)	7 (7.9)	4 (4.9)

[a]CS: combatting stress.

[b]WW: working with worry.

[c]BR: building resilience.

[d]PSS: Perceived Stress Scale.

[e]GAD-7: Generalized Anxiety Disorder-7.

[f]PHQ-8: Patient Health Questionnaire-8.

[g]BRS: Brief Resilience Scale.

XSL•FO

RenderX

Secondary Outcomes

ITT Analyses

Overview

The following are based on ITT analyses that include data from all randomized participants (except for those not meeting the eligibility criteria; see Multimedia Appendix 4 for PP results). EMMs for each secondary outcome (grouped by study arm and time point) are shown in Table 5. All study arms reported significant within-group improvements from t0 to t1 on all secondary outcomes (all $P<.05$), except for symptoms of depression (PHQ-8) and resilience (BRS) in the control group.

Table 5. Estimated marginal means (EMMs) from linear mixed-effects models at each of the 3 study time points shown for each intervention arm and each secondary outcome measure.

Outcome	Time point		
	Baseline (t0)	Postintervention (t1)	Follow-up (t2)
	EMM (SE; 95% CI)	EMM (SE; 95% CI)	EMM (SE; 95% CI)
PSS[a]			
Control	16.86 (0.76; 15.37-18.35)	15.19[b] (0.76; 13.69-16.69)	14.62[c] (0.76; 13.12-16.12)
CS[d]	16.35 (0.76; 14.86-17.84)	13.30[c] (0.77; 11.79-14.80)	12.59[c] (0.77; 11.08-14.10)
WW[e]	17.14 (0.75; 15.68-18.61)	12.93[c] (0.76; 11.44-14.42)	12.73[c] (0.76; 11.24-14.22)
BR[f]	17.76 (0.74; 16.29-19.22)	13.49[c] (0.76; 12.01-14.98)	13.17[c] (0.77; 11.66-14.68)
GAD-7[g]			
Control	6.38 (0.50; 5.41-7.36)	5.26[b] (0.50; 4.28-6.24)	5.42[h] (0.50; 4.44-6.40)
CS	6.20 (0.50; 5.23-7.18)	4.53[c] (0.50; 3.54-5.52)	4.01[c] (0.51; 3.02-5.00)
WW	6.57 (0.49; 5.61-7.53)	4.12[c] (0.50; 3.15-5.10)	4.12[c] (0.50; 3.15-5.10)
BR	7.02 (0.49; 6.06-7.98)	4.69[c] (0.50; 3.72-5.66)	4.42[c] (0.51; 3.43-5.41)
PHQ-8[i]			
Control	6.62 (0.51; 5.61-7.63)	6.38[j] (0.52; 5.37-7.39)	6.06[j] (0.52; 5.04-7.08)
CS	6.68 (0.51; 5.67-7.69)	4.94[c] (0.52; 3.92-5.96)	4.56[c] (0.52; 3.53-5.59)
WW	6.85 (0.51; 5.85-7.84)	4.36[c] (0.51; 3.35-5.36)	4.56[c] (0.51; 3.55-5.57)
BR	7.34 (0.50; 6.35-8.33)	5.53[c] (0.51; 4.53-6.54)	5.33[c] (0.52; 4.31-6.36)
BRS[k]			
Control	20.37 (0.52; 19.36-21.38)	20.90[j] (0.52; 19.89-21.92)	21.18[h] (0.52; 20.17-22.20)
CS	19.63 (0.52; 18.62-20.64)	21.03[c] (0.52; 20.01-22.05)	21.30[c] (0.52; 20.27-22.33)
WW	19.00 (0.51; 18.00-20.00)	20.70[c] (0.51; 19.70-21.71)	20.85[c] (0.51; 19.84-21.86)
BR	19.05 (0.50; 18.06-20.04)	20.52[c] (0.51; 19.51-21.52)	21.29[c] (0.52; 20.27-22.32)

[a]PSS: Perceived Stress Scale.

[b]$P\leq.01$; denotes significance of within-group contrasts comparing t0 to t1 and t0 to t2 for each outcome (P values are unadjusted).

[c]$P\leq.001$; denotes significance of within-group contrasts comparing t0 to t1 and t0 to t2 for each outcome (P values are unadjusted).

[d]CS: combatting stress.

[e]WW: working with worry.

[f]BR: building resilience.

[g]GAD-7: Generalized Anxiety Disorder-7.

[h]$P<.05$; denotes significance of within-group contrasts comparing t0 to t1 and t0 to t2 for each outcome (P values are unadjusted).

[i]PHQ-8: Patient Health Questionnaire-8.

[j]$P>.05$.

[k]BRS: Brief Resilience Scale.

Combatting Stress

Participants in the *combatting stress* arm reported a trend toward larger decreases in perceived stress (t1−t0: Hedges *g*=0.24, 95% CI −0.05 to 0.53; t2−t0: Hedges *g*=0.27, 95% CI −0.02 to 0.56) and larger increases in resilience (t1−t0: Hedges *g*=0.24, 95% CI −0.05 to 0.53; t2−t0: Hedges *g*=0.24, 95% CI −0.05 to 0.53) at both time points; however, none reached significance at an α level of .05 in this sample. Participants also reported greater improvement in symptoms of depression at t1 and t2 (t1−t0*:* Hedges *g*=0.37, 95% CI 0.08-0.66, *P*=.01; t2−t0: Hedges *g*=0.38, 95% CI 0.09-0.67, *P*=.009) and in symptoms of anxiety at t2 (Hedges *g*=0.30, 95% CI 0.01-0.59; *P*=.04) than the control group (Table 6).

Table 6. Contrasts and between-group (intervention versus control) effect size calculations from linear mixed-effects models (LMMs) applied to each secondary outcome for both intention-to-treat (ITT) and subgroup analyses.

Outcome	Estimate[a] (SE; 95% CI)	P value		Hedges g (95% CI)	
		LMM[b]	Tukey adjusted[c]	ITT	Subgroup
PSS[d,e]					
t1[f] minus t0[g]					
CS[h]	1.38 (0.83; −0.57 to 3.33)	.10	.34	0.24 (−0.05 to 0.53)	0.29 (−0.08 to 0.66)
WW[i]	2.54 (0.82; 0.93 to 4.16)	.002	.01	0.45 (0.16 to 0.74)	0.66 (0.28 to 1.04)
BR[j]	2.59 (0.82; 0.97 to 4.21)	.002	.009	0.45 (0.17 to 0.74)	0.47 (0.10 to 0.84)
All[k]	2.18 (0.67; 0.86 to 3.50)	.001	.004	0.38 (0.18 to 0.58)	0.47 (0.21 to 0.73)
t2[l] minus t0					
CS	1.52 (0.83; −0.11 to 3.16)	.07	.26	0.27 (−0.02 to 0.56)	0.41 (0.04 to 0.78)
WW	2.18 (0.83; 0.55 to 3.80)	.009	.04	0.38 (0.09 to 0.67)	0.64 (0.26 to 1.02)
BR	2.34 (0.84; 0.70 to 3.98)	.005	.03	0.40 (0.12 to 0.69)	0.53 (0.16 to 0.90)
All	2.02 (0.68; 0.69 to 3.34)	.003	.008	0.35 (0.15 to 0.55)	0.52 (0.26 to 0.78)
GAD-7[m,n]					
t0 minus t1					
CS	0.55 (0.59; −0.61 to 1.71)	.35	.79	0.14 (−0.15 to 0.42)	0.29 (−0.09 to 0.66)
WW	1.32 (0.59; 0.17 to 2.48)	.02	.11	0.33 (0.04 to 0.61)	0.46 (0.09 to 0.84)
BR	1.21 (0.59; 0.06 to 2.36)	.04	.17	0.30 (0.01 to 0.58)	0.41 (0.04 to 0.78)
All	1.03 (0.48; 0.09 to 1.97)	.03	.08	0.25 (0.05 to 0.46)	0.39 (0.12 to 0.65)
t0 minus t2					
CS	1.23 (0.59; 0.06 to 2.40)	.04	.16	0.30 (0.01 to 0.59)	0.49 (0.11 to 0.88)
WW	1.48 (0.59; 0.33 to 2.64)	.01	.06	0.36 (0.08 to 0.65)	0.64 (0.25 to 1.02)
BR	1.64 (0.60; 0.47 to 2.81)	.006	.03	0.40 (0.11 to 0.68)	0.58 (0.20 to 0.96)
All	1.45 (0.48; 0.51 to 2.40)	.003	.008	0.35 (0.15 to 0.56)	0.57 (0.30 to 0.84)
PHQ-8[o,p]					
t0 minus t1					
CS	1.51 (0.59; 0.34 to 2.67)	.01	.06	0.37 (0.08 to 0.66)	0.47 (0.08 to 0.85)
WW	2.25 (0.59; 1.09 to 3.41)	<.001	.001	0.55 (0.26 to 0.84)	0.70 (0.32 to 1.09)
BR	1.57 (0.59; 0.41 to 2.73)	.008	.04	0.38 (0.10 to 0.67)	0.46 (0.08 to 0.83)
All	1.78 (0.48; 0.83 to 2.72)	<.001	<.001	0.43 (0.23 to 0.64)	0.54 (0.28 to 0.81)
t0 minus t2					
CS	1.56 (0.60; 0.39 to 2.74)	.009	.045	0.38 (0.09 to 0.67)	0.57 (0.19 to 0.96)
WW	1.72 (0.59; 0.56 to 2.89)	.004	.02	0.42 (0.13 to 0.71)	0.62 (0.24 to 1.00)
BR	1.44 (0.60; 0.27 to 2.62)	.02	.08	0.35 (0.06 to 0.63)	0.46 (0.09 to 0.84)
All	1.58 (0.49; 0.63 to 2.53)	.001	.003	0.38 (0.18 to 0.59)	0.55 (0.28 to 0.82)
BRS[q,r]					
t0 minus t1					
CS	0.87 (0.52; −0.15 to 1.90)	.10	.34	0.24 (−0.05 to 0.53)	0.35 (0.16 to 0.85)
WW	1.18 (0.52; 0.16 to 2.19)	.02	.11	0.33 (0.04 to 0.62)	0.66 (0.19 to 1.13)
BR	0.94 (0.52; −0.08 to 1.96)	.07	.27	0.26 (−0.02 to 0.55)	0.44 (−0.06 to 0.93)

Outcome	Estimate[a] (SE; 95% CI)	P value		Hedges g (95% CI)	
		LMM[b]	Tukey adjusted[c]	ITT	Subgroup
All	1.00 (0.42; 0.17 to 1.83)	.02	.049	0.28 (0.08 to 0.48)	0.50 (0.16 to 0.83)
t0 minus t2					
CS	0.86 (0.53; −0.17 to 1.89)	.10	.36	0.24 (−0.05 to 0.53)	0.18 (−0.32 to 0.68)
WW	1.04 (0.52; 0.01 to 2.06)	.047	.19	0.29 (0.00 to 0.58)	0.47 (0.00 to 0.94)
BR	1.43 (0.53; 0.40 to 2.47)	.006	.03	0.39 (0.11 to 0.68)	0.38 (−0.11 to 0.87)
All	1.11 (0.43; 0.27 to 1.94)	.01	.03	0.30 (0.10 to 0.51)	0.36 (0.02 to 0.69)

[a]Contrast estimates from LMMs.

[b]P value from LMM (group×time interaction terms).

[c]P value following Tukey adjustment for all pairwise comparisons (only comparisons of interest are shown).

[d]PSS: Perceived Stress Scale.

[e]Sample size for ITT analyses n=383; sample size for subgroup analyses n=236.

[f]t1: time point 1.

[g]t0: time point 0.

[h]CS: combatting stress.

[i]WW: working with worry.

[j]BR: building resilience.

[k]Pooled effect of all intervention arms.

[l]t2: time point 2.

[m]GAD-7: Generalized Anxiety Disorder-7.

[n]Sample size for ITT analyses n=383; sample size for subgroup analyses n=227.

[o]PHQ-8: Patient Health Questionnaire-8.

[p]Sample size for ITT analyses n=383; sample size for subgroup analyses n=229.

[q]BRS: Brief Resilience Scale.

[r]Sample size for ITT analyses n=383; sample size for subgroup analyses n=142.

Working With Worry

Participants in the *working with worry* arm reported greater improvements in symptoms of anxiety at t1 and t2 than the control group (t1 minus t0: Hedges g=0.33, 95% CI 0.04-0.61, P=.02; t2 minus t0: Hedges g=0.36, 95% CI 0.08-0.65, P=.01), as well as greater improvements across all other secondary outcomes (all P<.05; Hedges g range 0.33-0.55; Table 6). All improvements were maintained at t2 (all P<.05; Hedges g range 0.29-0.42). Effect sizes were largest for symptoms of depression at both t1 (Hedges g=0.55, 95% CI 0.26-0.84) and t2 (Hedges g=0.42, 95% CI 0.13-0.71).

Building Resilience

Participants in the *building resilience* arm reported a trend toward larger increases in resilience at t1 than the control group (t1 minus t0: Hedges g=0.26, 95% CI −0.02 to 0.55), which emerged as significant at t2 (t2 minus t0: Hedges g=0.39, 95% CI 0.11-0.68; P=.006). In addition, participants reported significantly greater improvements across all other secondary outcomes at t1 (all P<.05; Hedges g range 0.30-0.45), which were maintained at t2 (all P<.05; Hedges g range 0.35-0.40; Table 6).

A comparison of the overall (pooled) effect of all intervention arms relative to the control group revealed significantly greater improvement for all 4 secondary outcome measures (Hedges g range 0.25–0.43; Table 6). In addition, post hoc contrasts on the LMM estimates suggested that none of the intervention arms were significantly different from one another when comparing t0 to t1 or t2 for any of the secondary outcome measures (all P>.05), although this study was not powered to detect such differences. Finally, when comparing score changes from t1 to t2, there were no significant differences between the control group and any of the intervention arms for any of the secondary outcome measures (all P>.05).

Subgroup Analyses

Subgroup analyses were performed for all secondary outcome measures to explore intervention effects for participants self-reporting at least mild symptoms (or moderately low resilience) at t0 (see Multimedia Appendix 7 for subgroup sample sizes and baseline scores). Findings were largely equivalent to the ITT analysis, although between-group effect sizes were generally larger, ranging from 0.39 (95% CI 0.12-0.65) to 0.54 (95% CI 0.28-0.81) when pooling the intervention arms (Table 6).

Exploratory Analyses

Exploratory multiple logistic regression suggested that participants working in health care (b=−2.11, SE 0.86; P=.01), finance, insurance, or real estate (b=−2.57, SE 0.88; P=.004), and professional services (b=−1.87, SE 0.91; P=.04) were less likely to complete their allocated intervention relative to those working in industrials (the reference category). The completion rate for industrials was 83.3% compared with 51.9% for health

care, 41.2% for finance, insurance, or real estate, and 55.6% for professional services.

In addition, participants with higher PHQ-8 scores at baseline were less likely to complete their intervention (b=−0.17, SE 0.05; $P<.001$). None of the demographic variables collected in this study predicted completion (all $P>.05$). These analyses also confirmed that participants allocated to *combatting stress* were more likely to complete their intervention than those allocated to *building resilience* (b=−1.05, SE 0.37; $P=.005$) while controlling for all other baseline variables.

Discussion

Principal Findings

Intervention research can be undermined by problems with intervention delivery, acceptability, participant retention, and smaller than anticipated effect sizes. Therefore, guidelines suggest conducting pilot studies to test trial feasibility and estimate important trial parameters before running a definitive trial [61]. This study reports on the feasibility and preliminary efficacy of 3 interventions featured on the *Unmind* MHapp when delivered to working adults in the United Kingdom. The study methods and interventions were found to be feasible, and all preregistered progression criteria were met. This suggests that a definitive trial is warranted, although several minor proposed protocol amendments are discussed.

Participant retention and intervention adherence were largely consistent with or higher than those in comparable studies. For instance, only 7% (27/383) of participants were lost to attrition at follow-up, which compares favorably with recent meta-analyses reporting average attrition rates of 23% to 48% for MHapp trials [11,16,62]. This may be because of the use of the Prolific recruitment platform, which is associated with high participant response rates [35], and reimbursing participants at each study assessment. It may also suggest that participants, on average, perceived the *Unmind* platform as helpful and engaging and were thus motivated to complete the study.

Objective engagement data suggested that 67.8% (196/289) of randomized participants (and 196/237, 82.7% of those starting their intervention) completed all intervention sessions, which is similar to or higher than average completion rates ranging between 30% and 65% for other MHapp interventions [11,16,53,63]. These engagement rates are particularly encouraging, considering that participants were not recruited on the basis of seeking help for a mental health problem and were not randomized to an intervention based on scoring poorly on the target outcome at baseline.

Despite these high levels of engagement, the study interventions were brief, comprising approximately 60 to 80 minutes of learning over 7 sessions. In addition, the study used a nonclinical sample, and regression analyses suggested that participants with higher symptom severity at baseline were less likely to complete their allocated intervention. Thus, these findings may not be generalizable to other interventions featured on the *Unmind* app or to other study populations. Future studies designed to evaluate the use of the *Unmind* app in subclinical populations are currently being planned. Interestingly, although participants

with higher baseline symptoms were less likely to complete their intervention, feedback ratings at postintervention were largely equivalent across the study sample. For example, 87.1% (203/233) of all participants reported being either *satisfied* or *very satisfied* with their intervention and rated the quality of their intervention as either *good* or *excellent*, and these ratings did not differ for participants with more severe symptoms at baseline.

Although all 3 interventions met progression criteria, participants randomized to the *building resilience* arm were marginally less likely to start their allocated intervention and complete all intervention sessions after starting. A potential explanation is that participants may have felt less motivated by the theme of resilience as compared with stress and anxiety. In addition, participants in the *building resilience* arm reported marginally worse mental health scores at baseline, which may have negatively affected engagement. Although participant feedback was largely equivalent across the 3 interventions, there were several marginal (not statistically significant) differences that may be a contributing factor. For example, compared with *combatting stress*, participants in the *building resilience* arm were slightly less likely to rate the intervention as *good* relative to *okay* or *poor*. Although it is difficult to draw firm conclusions from these data, it will be important to test whether such differences persist in a definitive trial and the extent to which any changes or improvements to the *building resilience* intervention are warranted.

The findings from this pilot study revealed several opportunities for minor protocol improvements before running a definitive trial. First, of the 383 participants, 11 (2.9%) reported not fully understanding the study instructions, and 1.7% (5/289) of participants randomized to an intervention engaged with the wrong intervention. This could be addressed by providing participants with detailed video instructions and implementing a brief quiz to ensure that all participants understand how to access their allocated intervention. If feasible, participants could be given access to a modified, study-specific version of the *Unmind* platform that only includes the interventions being tested. Second, discrepancies between self-reported and objective in-app engagement meant that 48% (45/93) of participants who stopped using the *Unmind* app did not provide feedback on their reasons for discontinuing their intervention. This could be addressed by restructuring the feedback questionnaire to capture data from all participants, regardless of self-reported engagement. Third, 0.9% (2/233) of participants reported experiencing negative effects while completing their intervention; however, qualitative feedback suggested that both instances referred to frustration with the intervention (not understanding the content or not finding it useful). This question could be modified to capture *lasting* bad effects (refer to the study by Crawford et al [64]) with more specific clarifying questions to better differentiate harmful effects from frustration and/or lack of intervention enjoyment. Finally, 4% (16/400) of participants were excluded from the analysis as they reported being unemployed at baseline, which was in contrast to their prescreening responses. Although the Prolific platform precludes the use of additional screening questions at baseline, the sample

size calculation for a definitive trial could be adjusted to account for this potential discrepancy.

Although the study was not powered for formal hypothesis testing, preliminary efficacy findings suggested that the study interventions were associated with small to moderate between-group improvements in ≥1 mental health outcomes, which were maintained at the 1-month follow-up. This is consistent with findings from meta-analyses of previous MHapp trials [11,12,26]. All 3 interventions resulted in larger improvements in symptoms of depression, anxiety, perceived stress, and resilience than the control group, although smaller effects (Hedges g<0.3) did not reach statistical significance. An appropriately powered definitive trial may be more likely to capture such small effects with greater precision and, where possible, should aim to report on the clinical importance of these findings (eg, with regard to minimal important difference thresholds). These efficacy findings were consistent across both PP and ITT analyses and are particularly promising, given the brevity of the interventions. In addition, the findings were robust across subgroup analyses that only included participants with at least mild problems at baseline. However, 2 further patterns emerged that merit discussion.

First, relative to baseline, participants in the control group reported statistically significant improvements in stress and anxiety at both study time points and improvements in resilience at the 1-month follow-up, despite not having access to any study interventions. Although this may reflect phenomena such as practice effects or regression to the mean, it is worth noting that baseline data were collected several weeks after the commencement of a third national UK lockdown (in response to rising cases of SARS-CoV-2), whereas follow-up data were collected after the start of a phased easing of restrictions. A longitudinal survey conducted in England suggests that symptoms of anxiety and depression tend to rise rapidly during the early stages of a lockdown but decline quickly thereafter [65], which may partly explain the changes in mental health scores reported by the control group. Thus, the present efficacy findings may not be directly generalizable to other contexts. Importantly, the intervention groups reported larger improvements in mental health despite significant changes in the control group, and MHapps such as *Unmind* may be an effective way of delivering accessible mental health support to workforces working remotely or during times of national crisis.

Second, although all 3 interventions resulted in significant improvements for at least one mental health outcome, the study interventions did not appear to be sensitive or specific to their target outcome. For example, the *combatting stress* arm resulted in small between-group (intervention vs control) improvements in perceived stress (not statistically significant), as well as significant reductions in symptoms of depression. Similarly, participants in the *working with worry* arm reported significant improvements across all outcomes, despite the intervention specifically targeting anxiety. A potential explanation for this is that different facets of mental health tend to strongly covary with one another [66], and transdiagnostic research suggests that symptoms of different mental health problems often respond to the same treatments [67]. The present findings lend credence to this, as the study interventions all primarily involve similar CBT-based techniques (identifying and challenging negative thinking patterns, problem solving, and breaking negative cycles of thinking or behaviors). As this study was not powered to detect differences between intervention arms (or to test whether effect sizes for some outcomes were significantly different from others), it is not possible to draw meaningful conclusions from these data. Future studies should aim to test for any such differences so that users of the *Unmind* app can access interventions that are most likely to benefit their individual problem areas.

Strengths and Limitations

This study had several strengths. First, intervention adherence and engagement were objectively captured via *Unmind* (the intervention platform). This is important, given the recent evidence of substantial discrepancies between self-reported and objective adherence in digital interventions [68]. Second, participant retention was very high, which is extremely important as missing data can reduce statistical power and lead to biased intervention effects. Second, all study outcomes and analyses were prospectively registered, precluding any selective reporting of outcomes or nonpublication of findings [69]. Third, the study recruited a sample representative of the general UK population with respect to age, sex, and ethnicity, and participants were employed across a variety of industries. This is important as *Unmind* is designed for use across a broad range of demographics, and a lack of diversity within study samples can limit the generalizability of the study findings.

A limitation of this study is the use of a passive no-intervention control group, as opposed to an active control group in which participants engage with activities matched for duration, attention, and interest. Passive controls do not allow true intervention effects to be differentiated from nonspecific placebo effects and may introduce nocebo effects [31]. Although it is useful in practical terms to estimate this combined effect of the *Unmind* app, active controls are needed to fully understand the mechanisms underlying its effects. In addition, despite being randomly assigned, participants in the control group had slightly higher levels of self-reported resilience at baseline, and it is unknown whether this may have affected between-group differences in resilience scores over time. In addition, as with most web-based trials, participants were not blinded to group allocation. Finally, although the Prolific recruitment platform has several strengths, the participant pool was entirely self-selected, and it remains unknown to what extent the present findings are generalizable to working adults nationwide.

Conclusions

This study assessed the feasibility of conducting a future definitive RCT to evaluate the efficacy of 3 brief interventions featured on the *Unmind* MHapp. The study methods and interventions were found to be feasible, and all preregistered criteria for progression to a definitive trial were met. Several minor protocol amendments have been suggested. Preliminary efficacy findings indicate that the study interventions may result in improved mental health outcomes when offered to working adults.

XSL•FO

RenderX

Authors' Contributions

ME, HB, and KC conceived and designed the study. ME, HB, and RM collected the data. ME analyzed the data and wrote the initial draft of the manuscript. All authors have discussed the results and contributed to the final manuscript.

Conflicts of Interest

This study was funded in full by Unmind Ltd, the creator of the intervention evaluated. The authors ME, RM, and HB are employed by and own share options at Unmind Ltd. Author KC is a formal academic collaborator and receives financial compensation to provide scientific and research consulting services to Unmind Ltd.

Multimedia Appendix 1
Postintervention feedback questions.
[DOCX File , 22 KB - formative_v6i3e34032_app1.docx]

Multimedia Appendix 2
Deviations from the preregistered data analysis plan.
[DOCX File , 23 KB - formative_v6i3e34032_app2.docx]

Multimedia Appendix 3
Reliable change index calculations.
[DOCX File , 21 KB - formative_v6i3e34032_app3.docx]

Multimedia Appendix 4
Per-protocol analyses.
[DOCX File , 28 KB - formative_v6i3e34032_app4.docx]

Multimedia Appendix 5
Postintervention feedback ratings for subgroups.
[DOCX File , 28 KB - formative_v6i3e34032_app5.docx]

Multimedia Appendix 6
Participants' self-reported reasons for not starting or discontinuing their intervention.
[DOCX File , 23 KB - formative_v6i3e34032_app6.docx]

Multimedia Appendix 7
Subgroup sample sizes and baseline scores.
[DOCX File , 23 KB - formative_v6i3e34032_app7.docx]

Multimedia Appendix 8
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1135 KB - formative_v6i3e34032_app8.pdf]

References

1. Pinheiro M, Ivandic I, Razzouk D. The economic impact of mental disorders and mental health problems in the workplace. In: Razzouk D, editor. Mental Health Economics: The Costs and Benefits of Psychiatric Care. Cham: Springer International Publishing; 2017:415-430.
2. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. Lancet 2013;382(9904):1575-1586. [doi: 10.1016/S0140-6736(13)61611-6] [Medline: 23993280]
3. Hampson E, Jacob A. Mental health and employers: refreshing the case for investment. Deloitte. 2020. URL: https://www2.deloitte.com/uk/en/pages/consulting/articles/mental-health-and-employers-refreshing-the-case-for-investment.html [accessed 2020-08-21]
4. Patel V, Maj M, Flisher A, De Silva MJ, Koschorke M, Prince M, WPA Zonal and Member Society Representatives. Reducing the treatment gap for mental disorders: a WPA survey. World Psychiatry 2010;9(3):169-176 [FREE Full text] [doi: 10.1002/j.2051-5545.2010.tb00305.x] [Medline: 20975864]

5. Thornicroft G, Chatterji S, Evans-Lacko S, Gruber M, Sampson N, Aguilar-Gaxiola S, et al. Undertreatment of people with major depressive disorder in 21 countries. Br J Psychiatry 2017;210(2):119-124 [FREE Full text] [doi: 10.1192/bjp.bp.116.188078] [Medline: 27908899]

6. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]

7. Bauer M, Glenn T, Geddes J, Gitlin M, Grof P, Kessing LV, et al. Smartphones in mental health: a critical review of background issues, current status and future concerns. Int J Bipolar Disord 2020;8(1):2 [FREE Full text] [doi: 10.1186/s40345-019-0164-x] [Medline: 31919635]

8. Deady M, Glozier N, Calvo R, Johnston D, Mackinnon A, Milne D, et al. Preventing depression using a smartphone app: a randomized controlled trial. Psychol Med 2020:1-10. [doi: 10.1017/S0033291720002081] [Medline: 32624013]

9. Tan L, Wang MJ, Modini M, Joyce S, Mykletun A, Christensen H, et al. Preventing the development of depression at work: a systematic review and meta-analysis of universal interventions in the workplace. BMC Med 2014;12:74 [FREE Full text] [doi: 10.1186/1741-7015-12-74] [Medline: 24886246]

10. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]

11. Carolan S, Harris PR, Cavanagh K. Improving employee well-being and effectiveness: systematic review and meta-analysis of Web-based psychological interventions delivered in the workplace. J Med Internet Res 2017;19(7):e271 [FREE Full text] [doi: 10.2196/jmir.7583] [Medline: 28747293]

12. Phillips EA, Gordeev VS, Schreyögg J. Effectiveness of occupational e-mental health interventions: a systematic review and meta-analysis of randomized controlled trials. Scand J Work Environ Health 2019;45(6):560-576 [FREE Full text] [doi: 10.5271/sjweh.3839] [Medline: 31184758]

13. Lau N, O'Daffer A, Colt S, Yi-Frazier JP, Palermo TM, McCauley E, et al. Android and iPhone mobile apps for psychosocial wellness and stress management: systematic search in app stores and literature review. JMIR Mhealth Uhealth 2020;8(5):e17798 [FREE Full text] [doi: 10.2196/17798] [Medline: 32357125]

14. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]

15. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]

16. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]

17. Beck AT. Cognitive Therapy and the Emotional Disorders. New York: International Universities Press; 1976.

18. Kabat-Zinn J. Wherever You Go, There You Are: Mindfulness Meditation in Everyday Life. Paris: Hachette Books; 1994.

19. Jacobson NS, Martell CR, Dimidjian S. Behavioral activation treatment for depression: returning to contextual roots. Clin Psychol Sci Pract 2001;8(3):255-270. [doi: 10.1093/clipsy.8.3.255]

20. Hayes SC, Strosahl KD, Wilson KG. Acceptance and Commitment Therapy: An Experiential Approach to Behavior Change. New York: Guilford Press; 1999.

21. Seligman ME, Csikszentmihalyi M. Positive psychology: an introduction. Am Psychol 2000;55(1):5-14. [doi: 10.1037/0003-066X.55.1.5]

22. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS Consensus Group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Br Med J 2016;355:i5239 [FREE Full text] [doi: 10.1136/bmj.i5239] [Medline: 27777223]

23. Eldridge SM, Lancaster GA, Campbell MJ, Thabane L, Hopewell S, Coleman CL, et al. Defining feasibility and pilot studies in preparation for randomised controlled trials: development of a conceptual framework. PLoS One 2016;11(3):e0150205 [FREE Full text] [doi: 10.1371/journal.pone.0150205] [Medline: 26978655]

24. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The benefit of web- and computer-based interventions for stress: a systematic review and meta-analysis. J Med Internet Res 2017;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]

25. Mewton L, Smith J, Rossouw P, Andrews G. Current perspectives on Internet-delivered cognitive behavioral therapy for adults with anxiety and related disorders. Psychol Res Behav Manag 2014;7:37-46 [FREE Full text] [doi: 10.2147/PRBM.S40879] [Medline: 24511246]

26. Stratton E, Lampit A, Choi I, Calvo RA, Harvey SB, Glozier N. Effectiveness of eHealth interventions for reducing mental health conditions in employees: a systematic review and meta-analysis. PLoS One 2017;12(12):e0189904 [FREE Full text] [doi: 10.1371/journal.pone.0189904] [Medline: 29267334]

27. Shrivastava A, Desousa A. Resilience: a psychobiological construct for psychiatric disorders. Indian J Psychiatry 2016;58(1):38-43 [FREE Full text] [doi: 10.4103/0019-5545.174365] [Medline: 26985103]

28. Kalin NH. The critical relationship between anxiety and depression. Am J Psychiatry 2020;177(5):365-367. [doi: 10.1176/appi.ajp.2020.20030305] [Medline: 32354270]

29. Khan S, Khan RA. Chronic stress leads to anxiety and depression. Ann Psychiatry Ment Health 2017;5(1):1091 [FREE Full text]

30. Parmar MK, Carpenter J, Sydes MR. More multiarm randomised trials of superiority are needed. Lancet 2014;384(9940):283-284. [doi: 10.1016/S0140-6736(14)61122-3] [Medline: 25066148]

31. Furukawa TA, Noma H, Caldwell DM, Honyashiki M, Shinohara K, Imai H, et al. Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. Acta Psychiatr Scand 2014;130(3):181-192. [doi: 10.1111/acps.12275] [Medline: 24697518]

32. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Br Med J 2010;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]

33. Juszczak E, Altman DG, Hopewell S, Schulz K. Reporting of multi-arm parallel-group randomized trials: extension of the CONSORT 2010 statement. J Am Med Assoc 2019;321(16):1610-1620. [doi: 10.1001/jama.2019.3087] [Medline: 31012939]

34. Prolific. URL: https://www.prolific.co/ [accessed 2014-08-20]

35. Peer E, Brandimarte L, Samat S, Acquisti A. Beyond the Turk: alternative platforms for crowdsourcing behavioral research. J Exp Soc Psychol 2017;70:153-163. [doi: 10.1016/j.jesp.2017.01.006]

36. Qualtrics. URL: https://www.qualtrics.com [accessed 2021-09-20]

37. O'Cathain A, Croot L, Duncan E, Rousseau N, Sworn K, Turner KM, et al. Guidance on how to develop complex interventions to improve health and healthcare. BMJ Open 2019;9(8):e029954 [FREE Full text] [doi: 10.1136/bmjopen-2019-029954] [Medline: 31420394]

38. Duggan C, Parry G, McMurran M, Davidson K, Dennis J. The recording of adverse events from psychological treatments in clinical trials: evidence from a review of NIHR-funded trials. Trials 2014;15:335 [FREE Full text] [doi: 10.1186/1745-6215-15-335] [Medline: 25158932]

39. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]

40. Rozental A, Andersson G, Boettcher J, Ebert DD, Cuijpers P, Knaevelsrud C, et al. Consensus statement on defining and measuring negative effects of internet interventions. Internet Interv 2014;1(1):12-19. [doi: 10.1016/j.invent.2014.02.001]

41. Cohen S, Williamson G. Perceived stress in a probability sample of the United States. In: Spacapan S, Oskamp S, editors. The Social Psychology of Health: The Claremont Symposium on Applied Social Psychology. Thousand Oaks: Sage Publications; 1989.

42. Lee EH. Review of the psychometric evidence of the perceived stress scale. Asian Nurs Res (Korean Soc Nurs Sci) 2012;6(4):121-127 [FREE Full text] [doi: 10.1016/j.anr.2012.08.004] [Medline: 25031113]

43. Baer RA, Carmody J, Hunsinger M. Weekly change in mindfulness and perceived stress in a mindfulness-based stress reduction program. J Clin Psychol 2012;68(7):755-765. [doi: 10.1002/jclp.21865] [Medline: 22623334]

44. Cavanagh K, Churchard A, O'Hanlon P, Mundy T, Votolato P, Jones F, et al. A randomised controlled trial of a brief online mindfulness-based intervention in a non-clinical population: replication and extension. Mindfulness (N Y) 2018;9(4):1191-1205 [FREE Full text] [doi: 10.1007/s12671-017-0856-1] [Medline: 30100934]

45. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

46. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the generalized anxiety disorder screener (GAD-7) in the general population. Med Care 2008;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]

47. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

48. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]

49. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. Ann Fam Med 2010;8(4):348-353 [FREE Full text] [doi: 10.1370/afm.1139] [Medline: 20644190]

50. Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The brief resilience scale: assessing the ability to bounce back. Int J Behav Med 2008;15(3):194-200. [doi: 10.1080/10705500802222972] [Medline: 18696313]

51. Smith BW, Epstein EM, Ortiz JA, Christopher PJ, Tooley EM. The foundations of resilience: what are the critical resources for bouncing back from stress? In: Prince-Embury S, Saklofske DH, editors. Resilience in Children, Adolescents, and Adults: Translating Research into Practice. New York: Springer; 2013:167-187.

52.   Mbuagbaw L, Kosa SD, Lawson DO, Stalteri R, Olaiya OR, Alotaibi A, et al. The reporting of progression criteria in protocols of pilot trials designed to assess the feasibility of main trials is insufficient: a meta-epidemiological study. Pilot Feasibility Stud 2019;5:120 [FREE Full text] [doi: 10.1186/s40814-019-0500-z] [Medline: 31700654]

53.   Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]

54.   Teare MD, Dimairo M, Shephard N, Hayman A, Whitehead A, Walters SJ. Sample size requirements to estimate key design parameters from external pilot randomised controlled trials: a simulation study. Trials 2014;15:264 [FREE Full text] [doi: 10.1186/1745-6215-15-264] [Medline: 24993581]

55.   Cohen S, Janicki-Deverts D. Who's stressed? Distributions of psychological stress in the United States in probability samples from 1983, 2006, and 2009. J Appl Soc Psychol 2012;42(6):1320-1334. [doi: 10.1111/j.1559-1816.2012.00900.x]

56.   Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991;59(1):12-19. [doi: 10.1037//0022-006x.59.1.12] [Medline: 2002127]

57.   Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]

58.   Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev Camb Philos Soc 2007;82(4):591-605. [doi: 10.1111/j.1469-185X.2007.00027.x] [Medline: 17944619]

59.   Population estimates for the UK, England and Wales, Scotland and Northern Ireland: mid 2019. Office for National Statistics. 2020. URL: https://www.ons.gov.uk/ [accessed 2001-08-20]

60.   Lambert M. Presidential address: what we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. Psychother Res 2007;17(1):1-14. [doi: 10.1080/10503300601032506]

61.   Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. Br Med J 2008;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]

62.   Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]

63.   Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]

64.   Crawford MJ, Thana L, Farquharson L, Palmer L, Hancock E, Bassett P, et al. Patient experience of negative effects of psychological treatment: results of a national survey†. Br J Psychiatry 2016;208(3):260-265. [doi: 10.1192/bjp.bp.114.162628] [Medline: 26932486]

65.   Fancourt D, Steptoe A, Bu F. Trajectories of anxiety and depressive symptoms during enforced isolation due to COVID-19 in England: a longitudinal observational study. Lancet Psychiatry 2021;8(2):141-149 [FREE Full text] [doi: 10.1016/S2215-0366(20)30482-X] [Medline: 33308420]

66.   Plana-Ripoll O, Pedersen CB, Holtz Y, Benros ME, Dalsgaard S, de Jonge P, et al. Exploring comorbidity within mental disorders among a Danish national population. JAMA Psychiatry 2019;76(3):259-270 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.3658] [Medline: 30649197]

67.   Caspi A, Moffitt TE. All for one and one for all: mental disorders in one dimension. Am J Psychiatry 2018;175(9):831-844 [FREE Full text] [doi: 10.1176/appi.ajp.2018.17121383] [Medline: 29621902]

68.   Flett JA, Fletcher BD, Riordan BC, Patterson T, Hayne H, Conner TS. The peril of self-reported adherence in digital interventions: a brief example. Internet Interv 2019;18:100267 [FREE Full text] [doi: 10.1016/j.invent.2019.100267] [Medline: 31890620]

69.   Al-Durra M, Nolan RP, Seto E, Cafazzo JA, Eysenbach G. Nonpublication rates and characteristics of registered randomized clinical trials in digital health: cross-sectional analysis. J Med Internet Res 2018;20(12):e11924 [FREE Full text] [doi: 10.2196/11924] [Medline: 30485832]

Abbreviations

ACT: acceptance and commitment therapy
BRS: Brief Resilience Scale
CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
EMM: estimated marginal mean
GAD: generalized anxiety disorder
ITT: intention-to-treat
LMM: linear mixed-effects model
MHapp: mental health app
MM: mindfulness meditation
PHQ: Patient Health Questionnaire
PP: per-protocol

XSL•FO

RenderX

PSS: Perceived Stress Scale
RCT: randomized controlled trial
t0: time point 0
t1: time point 1
t2: time point 2

XSL•FO
RenderX

Original Paper

Digital Life Coaching During Stem Cell Transplantation: Development and Usability Study

Rahul Banerjee[1], MD; Chiung-Yu Huang[2], PhD; Lisa Dunn[1], BSN; Jennifer Knoche[1], BSN; Chloe Ryan[1], BSc; Kelly Brassil[3], PhD; Lindsey Jackson[3], MPH; Dhiren Patel[3], PharmD; Mimi Lo[4], PharmD; Shagun Arora[1], MD; Sandy W Wong[1], MD; Jeffrey Wolf[1], MD; Thomas Martin III[1], MD; Anand Dhruva[1], MD; Nina Shah[1], MD

[1]Division of Hematology/Oncology, Department of Medicine, University of California San Francisco, San Francisco, CA, United States

[2]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, United States

[3]Pack Health, Birmingham, AL, United States

[4]Division of Hematology/Oncology, Department of Pharmacy, University of California San Francisco, San Francisco, CA, United States

Corresponding Author:
Rahul Banerjee, MD
Division of Hematology/Oncology
Department of Medicine
University of California San Francisco
400 Parnassus Avenue
San Francisco, CA, 94158
United States
Phone: 1 415 353 8000
Fax: 1 415 353 8000
Email: rahul.banerjee.md@gmail.com

Abstract

Background: For patients with multiple myeloma receiving high-dose chemotherapy followed by autologous stem cell transplantation (SCT), acute life disruptions and symptom burden may lead to worsened quality of life (QOL) and increased emotional distress. Digital life coaching (DLC), whereby trained coaches deliver personalized well-being–related support via phone calls and SMS text messaging, has been shown to improve QOL among SCT survivors. However, DLC has not been investigated during the acute peri-SCT period, which is generally characterized by symptomatic exacerbations and 2-week hospitalizations.

Objective: We launched a single-arm pilot study to investigate the feasibility of patient engagement with DLC during this intensive period.

Methods: We approached English-speaking adult patients with multiple myeloma undergoing autologous SCT at our center. Enrolled patients received 16 weeks of virtual access to a life coach beginning on day −5 before SCT. Coaches used structured frameworks to help patients identify and overcome personal barriers to well-being. Patients chose the coaching topics and preferred communication styles. Our primary endpoint was ongoing DLC engagement, defined as bidirectional conversations occurring at least once every 4 weeks during the study period. Secondary endpoints were electronic patient-reported outcome assessments of QOL, distress, and sleep disturbances.

Results: Of the 20 patients who were screened, 17 (85%) chose to enroll and 15 (75%) underwent SCT as planned. Of these 15 patients (median age 65 years, range 50-81 years), 11 (73%) demonstrated ongoing DLC engagement. The median frequency of bidirectional conversations during the 3-month study period was once every 6.2 days (range 3.9-28 days). During index hospitalizations with median lengths of stay of 16 days (range 14-31 days), the median frequency of conversations was once every 5.3 days (range 2.7-15 days). Electronic patient-reported outcome assessments (94% adherence) demonstrated an expected QOL nadir during the second week after SCT. The prevalence of elevated distress was highest immediately before and after SCT, with 69% of patients exhibiting elevated distress on day −5 and on day +2.

Conclusions: DLC may be feasible for older patients during intensive hospital-based cancer treatments such as autologous SCT for multiple myeloma. The limitations of our study include small sample size, selection bias among enrolled patients, and heterogeneity in DLC use. Based on the positive results of this pilot study, a larger phase 2 randomized study of DLC during SCT is underway to investigate the efficacy of DLC with regard to patient well-being.

Introduction

Multiple myeloma (MM) is an incurable hematologic malignancy in older adults. Unlike in many other malignancies, upfront use of myeloablative chemotherapy followed by autologous stem cell transplantation (SCT) remains the standard of care for MM in eligible patients [1,2]. Autologous SCT is marked by acute symptomatic toxicities during the first 100 days after transplantation, such as fatigue, pain, and anorexia [3-6]. Sudden functional limitations, increased isolation, and nonrestful inpatient environments may also contribute to emotional distress. Patients must be monitored closely for other post-SCT toxicities and may, on average, spend over 30% of their days during this 3-month period either hospitalized or at clinic appointments [7,8]. Furthermore, most patients with MM who are employed prior to undergoing SCT are unable to return to work thereafter [9]. These factors, in turn, may lead to significant personal costs from transportation-related expenses and missed workplace productivity.

Specific manifestations of peri-SCT life disruptions may include worsened quality of life (QOL), elevated anxiety or emotional distress, and worsened sleep disturbances. These symptoms are particularly relevant for patients with MM who tend to be older and may have poorer baseline QOL than patients with other malignancies [10]. Previous studies, while limited by substantial heterogeneity in patient populations and survey-based inventories, suggest that well-being reaches its nadir 1-2 weeks after SCT before recovering in subsequent months [5,11-16]. Even so, long-term consequences of these short-term exacerbations may include persistent decreases in QOL, lowered posttransplant medication adherence, psychological comorbidities, reliance on potentially inappropriate medications such as benzodiazepines, and increased risk of hospital readmissions [17-23]. Several hospital-based interventions have thus been studied to target well-being during SCT, such as scheduled palliative care consultations, structured exercise programs, acupuncture, music therapy, and programmed room lighting [24-29]. However, these strategies may be limited by in-person provider availability in the inpatient setting or the need for "extra" clinic appointments in the outpatient setting.

Digital life coaching (DLC), whereby patients receive well-being–related support from trained coaches via bidirectional phone calls and SMS text messaging on their personal phones, may be able to address these limitations because of its virtual and location-agnostic nature. Life coaching, whereby trained coaches provide support and longitudinal accountability to empower patients to set and accomplish personal goals, has been effective in several ambulatory cancer populations [30-39]. DLC can reach patients both during and after their index SCT hospitalizations, which is a priority for patient-facing digital tools in this population [40]. While feasible among SCT survivors [41,42], DLC-type interventions have not been studied during the acute 100-day period immediately preceding and following SCT. Given the paucity of studies on the use of communication-based digital technologies among older adults [43], it is similarly unclear how acceptable DLC can be for patients with MM. We thus launched a pilot study to assess the feasibility of a DLC program for patients with MM actively undergoing SCT.

Methods

Study Design and Intervention

We launched a single-arm, pilot feasibility study of DLC among adult patients with MM undergoing nontandem autologous SCT at our center. English proficiency and mobile phone ownership were required for study participation; however, neither smartphone ownership nor mobile app installation were needed. There were no restrictions on time frames for pre-SCT stem cell collection or on agents used for stem cell mobilization. Based on a 68% rate of engagement in a prior study of DLC among SCT survivors and an assumption that DLC engagement below 33% would not merit further study [41], we enrolled 15 patients to exclude the possibility (with 1-sided α .05 and 90% power) of DLC engagement falling below this threshold. Patients who enrolled in the study but did not ultimately undergo autologous SCT were replaced. All patients provided informed consent prior to study enrollment. This study was approved by the University of California San Francisco Institutional Review Board (ClinicalTrials.gov NCT04432818).

Regarding the DLC platform itself, 2 life coaches employed by Pack Health [44] were paired with all patients enrolled in this study. Both coaches were certified by the National Board for Health and Wellness Coaching [45]. Coaches reached out to patients to coordinate their first coaching call beginning no earlier than day −5 before SCT. Patients then received 16 weeks of free unlimited access to their life coach by phone, SMS text messaging, or email. Coaches used structured frameworks longitudinally, including the Transtheoretical Model, Fogg Behavior Model, SMART (Specific, Measurable, Achievable, Relevant, and Time-based) Framework, and Pathways Thinking, to help patients identify and overcome personal barriers to well-being [46-53]. The content of the coaching curriculum was highly personalized to each patient at any given time and was not standardized per the study protocol. However, coaches did attempt to discuss several components of wellness (physical health, mental health, nutrition, exercise, sleep, and financial health) at least once during the study period. Coaches were not medical providers and were not licensed to address medical or psychiatric issues; as such, coaches were instructed to refer patients back to their SCT providers with any clinical concerns.

Patients could communicate bidirectionally with their coach by SMS text messaging, phone call, or email. This contact information was standardized for each coach; in other words, patients could add their coach to their list of phone contacts and communicate with them as they would communicate with loved ones. While coaches encouraged the use of weekly check-in phone calls, the actual cadence and communication methods of coaching were personalized to each patient based on their individual goals and preferences. Caregivers were allowed to join or participate in coaching as well, although all content was specifically geared toward patients themselves. Coaches organized phone calls and responded to messages during business hours for the DLC vendor (8 AM to 5 PM CST, corresponding to GMT −6); all enrolled patients were on PST (GMT −8). If patients did not respond to messages from their coaches, follow-up messages were sent no more frequently than 3 times per week. Patients were not contacted by coaches after the conclusion of the study period.

Data Collection

The primary objective of this study was to assess the rate of ongoing patient engagement with the DLC program using an intent-to-treat approach during the 16-week study period. We adopted the definition of feasibility used in a previous study of DLC among SCT survivors—patient-initiated engagement at least once with the DLC platform in at least 68% of patients over a 3-month period [41]—with 2 a priori modifications. First, we focused on bidirectional conversations (including phone calls lasting at least 1 minute) as examples of meaningful engagement even if initiated by the coach. Second, we adopted a stricter definition of feasibility as at least one bidirectional conversation every 4 weeks during the 16-week period rather than at least once over 3 months. Our rationale for this second modification was to assess the practicality of *ongoing* patient-coach conversations across a dynamic 100-day period including index hospitalizations as well as the initiation of post-SCT maintenance therapy.

The secondary objectives of this study were to explore the results of email-based electronic patient-reported outcome (ePRO) assessments measuring QOL, emotional distress, and sleep disturbances. ePROs were assessed in accordance with the Checklist for Reporting Results of Internet E-Surveys [54]. To measure QOL, we administered the 10-item Patient-Reported Outcome Measurement Information System (PROMIS) Global Health Scale instrument v1.2. To measure emotional distress, we administered the single-item National Comprehensive Cancer Network (NCCN) Distress Thermometer excluding the problem list. To measure sleep disturbances, we administered the 4-item PROMIS Sleep Disturbance Short Form 4a. These inventories, comprising 15 questions across 4 pages including a welcome page, were assessed through a secure web-based Research Electronic Data Capture (REDCap; Vanderbilt University) platform hosted on the study site's server. Patients received unique weblinks to complete ePRO assessments through emails sent from REDCap to their personal email addresses. These emails were sent at 12 discrete timepoints: weekly for the first

8 weeks of the study (day −5 to day +50) and then biweekly thereafter (day +51 to day +106). No ePRO questions were mandatory, no incentives were provided for ePRO assessment completion, and responses were not editable or rewritable after completion of each timepoint. The results of ePRO assessments were not shared with patients' coaches or clinical teams at any time.

Data Analysis and Statistical Considerations

For our primary endpoint of ongoing DLC engagement, we calculated SCT-relative dates for each bidirectional conversation within each of four 4-week study subperiods: day −5 to +22, day +23 to +50, day +51 to +78, and day +79 to +106. Based on our sample size of 15 patients, we defined feasibility as the presence of ongoing engagement in 10 or more patients (ie, at least 68% of patients) based on a previous study of DLC among SCT survivors [41].

For our secondary objectives involving ePRO assessments, we converted raw PROMIS inventory scores into T-scores to reflect a reference population with a mean score of 50 and SD of 10 as per their respective scoring manuals. We analyzed PROMIS inventory scores and Distress Thermometer scores at each timepoint descriptively using medians and ranges. Based on previous studies of patients with cancer, we defined clinically meaningful changes in PROMIS instruments as an increase or decrease of 5 or more points [55-57]. Specifically, we compared median T-scores at each timepoint to their baseline values. We defined worsened physical or mental QOL as a decrease of 5 or more points and worsened sleep as an increase of 5 or more points. We separately calculated the percentages of patients at each ePRO timepoint with elevated distress, defined as a Distress Thermometer score of 4 or higher [58].

Given our small sample size, we did not perform any longitudinal modeling of ePRO data. All analyses were performed using Stata (StataCorp) and R version 4.0.2 (R Foundation for Statistical Computing).

Results

Enrollment and Baseline Characteristics

Of the 20 patients approached between June 2020 and November 2020, 17 (85%) enrolled as outlined in Figure 1. However, 2 of these 17 (12%) patients did not undergo planned SCT for medical reasons and thus were replaced before DLC initiation. Of the 15/20 (75%) remaining patients, 2 (13%) dropped out of the study before coaching was initiated: one because she felt that the planned frequency of coaching would become too intense during SCT hospitalization and another because of a personal emergency (wildfire-related evacuations) that prevented her from coordinating a time to speak with her coach before arriving to the hospital to undergo SCT. Thus, 13/15 (87%) patients completed ePRO assessments beyond the baseline assessment and were evaluable for our secondary endpoints of QOL, emotional distress, and sleep disturbances.

Figure 1. Flow diagram of enrolled patients. Abbreviations: DLC, digital life coaching; MM, multiple myeloma; SCT, stem cell transplantation.
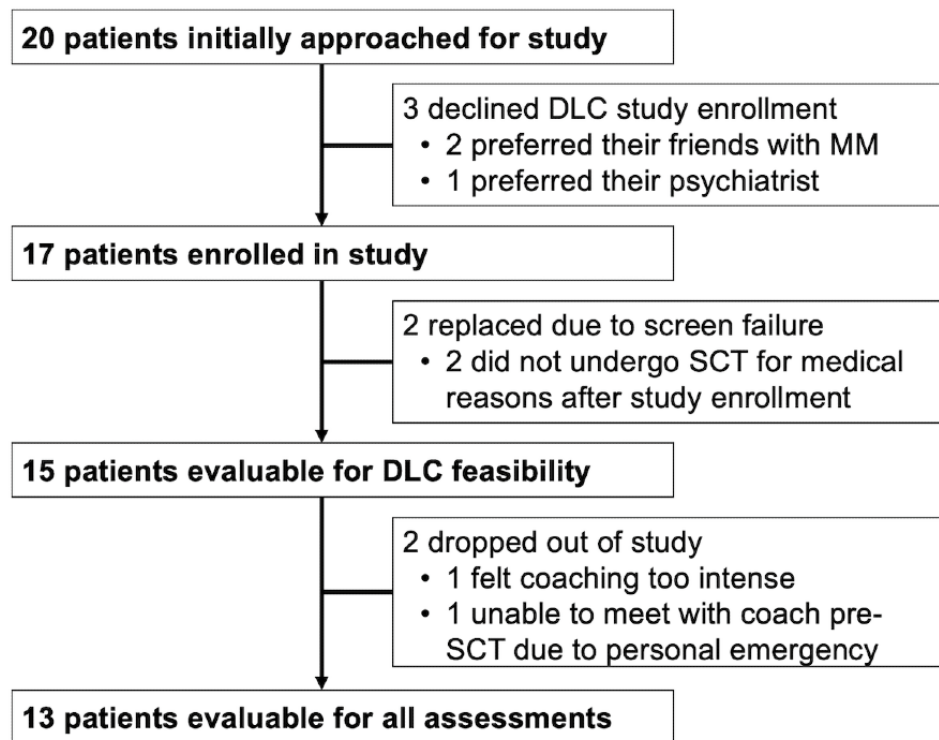
Baseline characteristics of the 15 patients evaluable for DLC feasibility are summarized in Table 1. The median age of the enrolled patients was 65 years (range 50-81 years); 4/15 (27%) patients were 70 years or older at the time of SCT. The median time between MM diagnosis and SCT was 7 months (range 4-108 months). A total of 4/15 patients (27%) had an Eastern Cooperative Oncology Group performance status of 1; the remainder had a performance status of 0. Most patients (12/15, 80%) received full-dose melphalan conditioning (200 mg/m^2) prior to SCT. All patients were hospitalized for SCT, with a median length of stay of 16 days (range 14-31 days).

Table 1. Baseline characteristics of evaluable patients (N=15).

Characteristic	Value, n (%)
Age at the time of SCT[a] (years)	
50-59.9	3 (20)
60-69.9	8 (53)
70-79.9	3 (20)
≥80	1 (7)
Gender	
Male	7 (47)
Female	8 (53)
Time since diagnosis (months)	
0-11.9	11 (73)
≥12	4 (27)
Performance status	
ECOG PS[b] 0	11 (73)
ECOG PS 1	4 (27)
Caregiver	
Spouse	13 (87)
Other[c]	2 (13)
Melphalan dose (mg/m^2)	
200	12 (80)
140	3 (20)

[a]SCT: stem cell transplantation.

[b]ECOG PS: Eastern Cooperative Oncology Group performance status.

[c]Other caregivers included an ex-spouse for one patient and a sibling for another patient.

Feasibility of the DLC Platform

Bidirectional conversations during the 16-week study period are depicted in Figure 2. Of the 15 enrolled patients who underwent SCT, 11 (73%) met our primary endpoint of ongoing engagement. Of the remaining 4/15 (27%) patients, 2 (50%) dropped out of the study prior to DLC initiation while an additional 2 (50%) demonstrated ongoing engagement only for the first 3 of the 4 study subperiods. For the 13/15 (87%) patients who received any coaching, the median number of conversation-days (defined as discrete days with at least one bidirectional conversation) during the 16-week study period was 18 (range 4-29). This corresponded to a median engagement frequency of 1 conversation every 6.2 days (range 3.9-28 days). During inpatient hospitalizations, this corresponding frequency was 1 conversation every 5.3 days (range 2.7-15 days). Of 240 conversation-days across 13 patients, 120 (50%) occurred via SMS text messaging while 109 (45%) occurred exclusively via phone calls. Of note, 69% (9/13) of patients never used emails to engage with their coaches.

We did not formally request feedback from patients about the DLC curriculum with respect to its feasibility or overlap with existing clinical resources. However, 2/13 (15%) patients did opt to reply at least once to automated ePRO assessment emails (correspondences that were then forwarded to the study team nonurgently). One patient wrote that her coach was "fantastic for answering questions, hearing and airing concerns…and mostly boosting hope, which is so very necessary in the MM world." A second patient's caregiver responded to an automated ePRO prompt inquiring about any additional medications for the management of neuropathy; however, whether this question had been redirected to the study team by the patient's coach was not specified.

Figure 2. Timeplot of bidirectional conversations. Days relative to SCT during which at least one bidirectional conversation took place are marked either by unboxed X icons (for outpatient) or boxed X icons (for inpatient). The 16-week study subperiod is divided into four 28-day subperiods as shown. *These patients underwent SCT but dropped out of the study prior to digital life coaching initiation. SCT: stem cell transplantation.
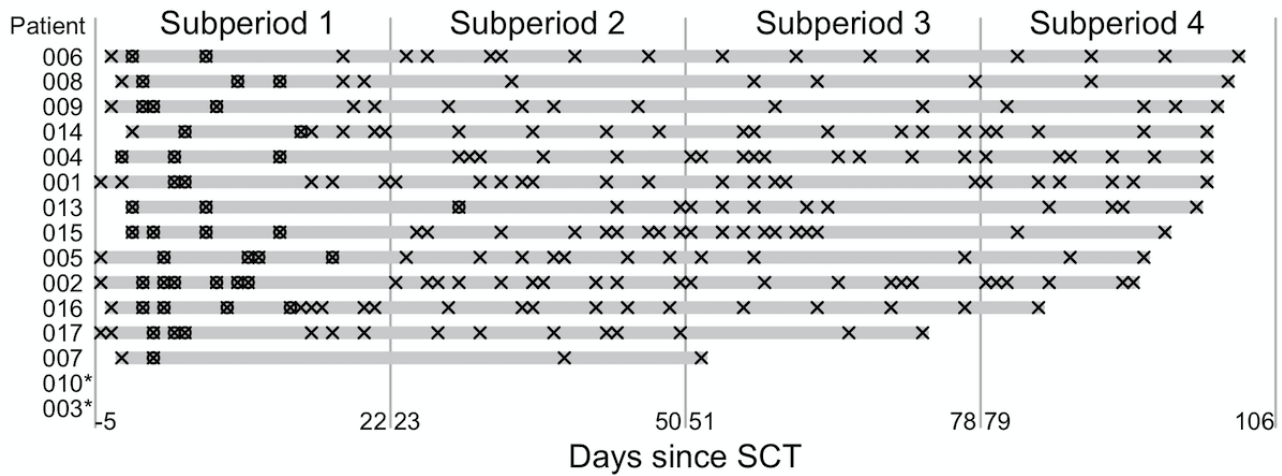
Longitudinal Patient-Reported Outcomes

The results of ePRO assessments for the 13/15 (87%) patients who received any coaching are depicted in Figure 3. A total of 94% (147/156) of ePRO assessments were completed, with a mean time of 3.3 minutes (range 1.1-17.9 minutes) spent per 15-question assessment. The results of these assessments were not shared with patients' coaches or clinical teams at any point. Compared to baseline values assessed at a median of day −5,

both the physical and mental components of QOL nadired during the second week after SCT (median day +9). There were no clinically meaningful exacerbations in sleep disturbances during the study period when compared to baseline. A total of 69% (9/13) of patients exhibited elevated distress at baseline (median day −5) and at the second timepoint (median day +2); this percentage decreased steadily in subsequent weeks to a nadir of 31% (4/13) in the eleventh week (median day +72).

Figure 3. Repeated ePRO assessments over time. For QOL (physical), QOL (mental), and sleep disturbances, values represent population-adjusted T-scores with a mean of 50 and SD of 10. Higher values represent better physical/mental QOL and worsened sleep, respectively. For % elevated distress, values represent the percentage of patients at each timepoint who reported elevated distress (defined as a Distress Thermometer score of 4 or higher) [58]. ePRO: electronic patient-reported outcome; PROMIS: Patient-Reported Outcome Measurement Information System; QOL: quality of life; SCT: stem cell transplantation.
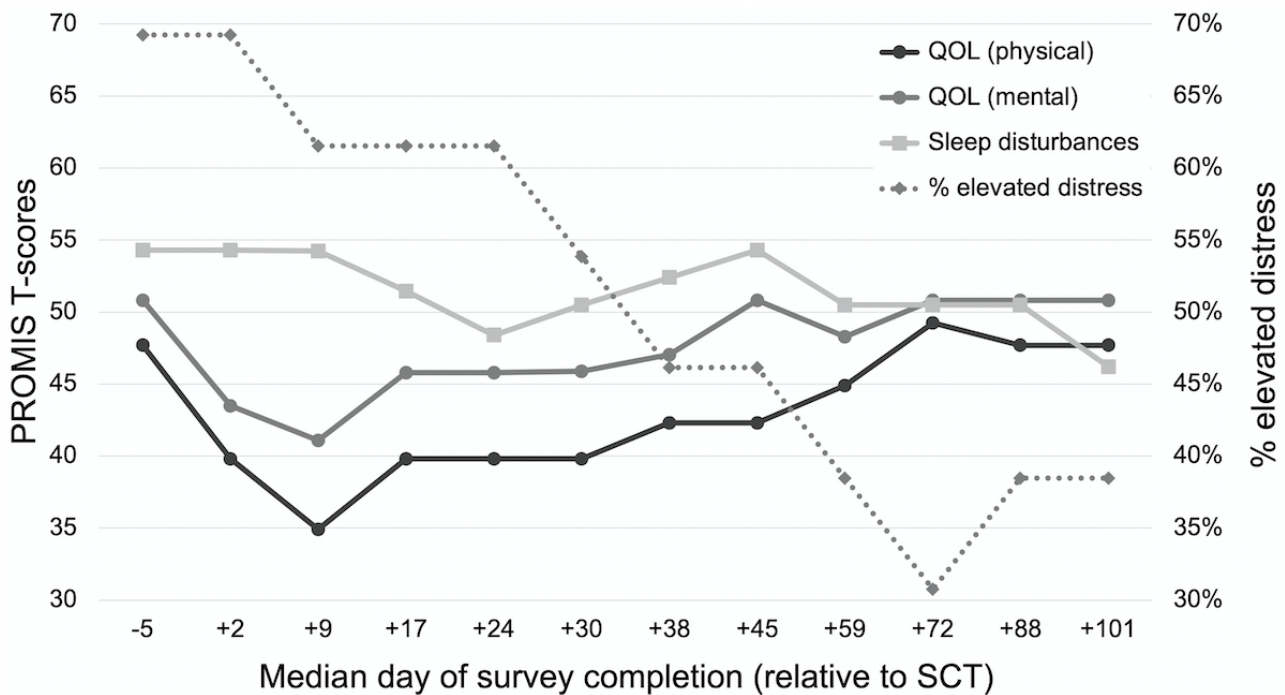
Discussion

Main Findings

In our single-arm pilot study, we found that DLC can be feasible for selected patients with MM during the intensive 100-day period encompassing autologous SCT. Although our study enrolled patients immediately prior to 2-week hospitalizations for myeloablative chemotherapy, the rate of ongoing patient engagement of 73% over a 3-month period was similar to or approached that of DLC-type interventions designed for ambulatory survivors beyond day +100 after SCT [41,42]. Hospitalized patients who underwent SCT in our study engaged bidirectionally with their coaches approximately once per week on average, a frequency comparable to that of unidirectional mobile health apps for patient education or ePRO completion [59,60]. Based on the positive results of this pilot study, a randomized phase 2 study of DLC versus usual care in this population is underway to investigate the efficacy of DLC (ClinicalTrials.gov NCT04589286).

One lesson learned from our study was the optimal timing of supportive interventions during intensive cancer-directed treatments. Our DLC intervention began on day −5 of SCT, which preceded hospital admission for high-dose chemotherapy by 2-3 days. Our rationale for this time frame was to focus on the acute posttransplant period itself, a time during which QOL is known to decrease because of increased symptom burden and acute life disruptions [11-15]. However, in contrast to 2 prior studies that demonstrated that distress peaked at count nadir or hospital discharge [3,15], this study found that elevated distress was highest at pre-SCT baseline. As a second observation of note, our narrow pre-SCT window also precluded participation for 1 of our 15 enrolled patients who was unable to connect with her coach before hospitalization. Earlier initiation of DLC, as we have implemented in our ongoing phase 2 study, may allow coaches to intervene during emotional distress when this symptom is at its peak while also improving the logistical adoptability of DLC for patients.

A second lesson learned from our study was the importance of flexibility and bidirectionality regarding how patients can engage with coaches. Our DLC platform allowed patients to call, text, or email their coaches in the same ways in which they might communicate with loved ones. All communications sent by coaches were intentionally worded to encourage subsequent responses from patients. Prior research has suggested that the promotion of interactivity may improve patients' perceptions of and patient engagement with digital platforms [61,62]. Although we did not formally test this hypothesis, these characteristics may have enhanced the "stickiness" of DLC (defined as the ease of continued use of a patient-facing platform over time) [63] during the peri-SCT period. As a counterexample within digital oncology, the randomized Southwest Oncology Group S1105 study of automated twice-weekly unidirectional SMS text messaging for women with breast cancer did not show a benefit in its primary endpoint of medication adherence [64]. The authors posited that their approach may not have sufficiently engaged patients to promote behavioral change.

Limitations and Future Directions

Our study nevertheless had several limitations; the most important one was the limited external validity from our small cohort of English-speaking patients who owned personal mobile phones. Although our intervention was deemed acceptable by patients who were planning to undergo SCT (with 85% of approached patients enrolling in the study), our patient population was relatively homogenous with regard to race, ethnicity, and marital status. Because racial and socioeconomic barriers affect SCT access and post-SCT outcomes in patients with MM [65,66], further investigation of this platform in a broader population of patients is needed. Two active areas of expansion for the DLC vendor include translation of content into other languages and development of an entirely landline-based curriculum (eg, using hospital-based phones and house-based phones) for patients who do not personally own mobile phones.

Other limitations of our study include our inability to isolate the specific coaching frameworks or themes that were most useful for patients. While patients communicated with their coaches approximately once per week on average, we did not directly gauge patient rationales for DLC engagement at each timepoint. As such, it is possible that patients may have maintained interactions with their coaches out of politeness rather than firm beliefs that the coaching was of value. Elucidating the specific impact of DLC on ePRO assessments of QOL is thus a key goal of our ongoing randomized phase 2 study. Furthermore, our ePRO results must be interpreted with caution given the small sample size and the presence of DLC itself as a potential confounder. Our finding that elevated distress was most prevalent in the days immediately before and after SCT was predicated on scores of 4 or higher on the single-question NCCN Distress Thermometer, a definition used previously in the SCT population [67,68]. It is unclear whether the use of longer survey-based instruments to assess distress, such as the Brief Symptom Inventory or Impact of Event Scale, would have yielded different results.

While DLC-type interventions have already been investigated in ambulatory cancer survivors [32,33,36-39,41,42], this is the first study to our knowledge to investigate DLC during a hospital-based cancer therapy such as SCT. DLC offers 3 possible benefits over traditional face-to-face tools during such intensive treatment modalities. First, although interventions with substantial in-person components may be more likely to improve distress in patients with cancer [69], additional in-person visits may be impractical for patients to attend in the setting of acute symptomatic toxicities. In contrast, DLC allows patients to access a centralized team of life coaches from the convenience of their phones regardless of their current location. Second, for patients who interact frequently with their coaches, "micro-learning" (a key functionality of mobile health tools) [70] may enhance the staying power of the coaching curriculum. Third, as discussed previously, DLC can readily accommodate individual patient preferences with regard to specific communication modalities and cadences.

However, whether DLC can truly improve the quality of supportive care during SCT requires further investigation in our

ongoing randomized phase 2 trial and subsequent investigations. Similarly, given a plethora of ePRO instruments used to assess QOL and distress in the hematopoietic stem cell transplant population [71], the PROMIS Global Health Scale and NCCN Distress Thermometer—both of which are relatively newer in this patient population—require further validation against longer survey-based instruments such as the Functional Assessment of Cancer Therapy Bone Marrow Transplantation or Brief Symptom Inventory assessments. Expansion of DLC into other transplantation settings, particularly for patients with acute leukemia undergoing allogeneic SCT, is warranted given that patients who undergo allogeneic SCT have a higher symptom burden than those who undergo autologous SCT [11,72]. This is an active area of investigation for our group.

Conclusions

Selected patients receiving high-dose chemotherapy followed by autologous SCT can engage meaningfully with life coaches using their phones, even during 2-week hospitalizations. A randomized phase 2 study to assess the efficacy of DLC in this population is underway. If future studies demonstrate the effectiveness of DLC in improving QOL and symptom burden during SCT, this type of intervention may eventually become a routine tool for supporting patient well-being during intensive cancer-directed therapies.

Conflicts of Interest

RB reports the following disclosures: Curio Science (honoraria), Guidepoint Global (consulting), Sanofi (consulting), SparkCures (consulting), and Pack Health (institutional research funding, not for this study). KB reports the following disclosures: Pack Health (employment), AbbVie (research), Astellas (research), BMS (research), Daiichi Sankyo (research), Genentech (research), GSK (research), and Sanofi (research). LJ reports the following disclosures: Pack Health (employment). DP reports the following disclosures: Pack Health (employment). ML reports the following disclosures: EUSA Pharma (consulting) and Oncopeptides (consulting). SWW reports the following disclosures: Amgen (consulting), BMS (research), Caelum (research), Genentech (research), Fortis (research), GSK (research), Janssen (research), and Sanofi (consulting). JW reports the following disclosures: Amgen (consulting), Celgene (consulting), Janssen (consulting), Novartis (consulting), and Takeda (consulting). TGM reports the following disclosures: Amgen (research), GSK (consulting), Janssen (research), Juno (consulting), Roche (consulting), Sanofi (research), and Seattle Genetics (research). NS reports the following disclosures: Amgen (consulting), Bluebird Bio (research), BMS/Celgene (consulting/research), CareDx (consulting), GSK (consulting), Indapta Therapeutics (consulting), Janssen (research), Karyopharm (consulting), Kite (consulting), Nektar (research), Poseida (research), Sanofi (consulting), Sutro Biopharma (research), and Teneobio (research). The remaining authors report no relevant disclosures.

References

1. Mikhael J, Ismaila N, Cheung MC, Costello C, Dhodapkar MV, Kumar S, et al. Treatment of multiple myeloma: ASCO and CCO Joint Clinical Practice Guideline. J Clin Oncol 2019 May 10;37(14):1228-1263. [doi: 10.1200/JCO.18.02096] [Medline: 30932732]

2. Dimopoulos M, Moreau P, Terpos E, Mateos M, Zweegman S, Cook G, EHA Guidelines Committee. Electronic address: guidelines@ehaweb.org, ESMO Guidelines Committee. Electronic address: clinicalguidelines@esmo.org. Multiple myeloma: EHA-ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2021 Mar;32(3):309-322. [doi: 10.1016/j.annonc.2020.11.014] [Medline: 33549387]

3. Anderson KO, Giralt SA, Mendoza TR, Brown JO, Neumann JL, Mobley GM, et al. Symptom burden in patients undergoing autologous stem-cell transplantation. Bone Marrow Transplant 2007 Jun;39(12):759-766. [doi: 10.1038/sj.bmt.1705664] [Medline: 17438588]

4. Wang XS, Shi Q, Shah ND, Heijnen CJ, Cohen EN, Reuben JM, et al. Inflammatory markers and development of symptom burden in patients with multiple myeloma during autologous stem cell transplantation. Clin Cancer Res 2014 Mar 01;20(5):1366-1374 [FREE Full text] [doi: 10.1158/1078-0432.CCR-13-2442] [Medline: 24423611]

5. Naegele M, Kirsch M, Ihorst G, Fierz K, Engelhardt M, De Geest S. Symptom experience of multiple myeloma (syMMex) patients treated with autologous stem cell transplantation following high-dose melphalan: a descriptive longitudinal study. Support Care Cancer 2018 Mar;26(3):833-841. [doi: 10.1007/s00520-017-3897-z] [Medline: 28918604]

6. Stamm SL, Spichiger E, Pabst T, Bachnick S, Jeitziner M. Symptom prevalence and health-related quality of life in patients undergoing autologous stem cell transplantation - A longitudinal observational study. Eur J Oncol Nurs 2021 Aug;53:101997 [FREE Full text] [doi: 10.1016/j.ejon.2021.101997] [Medline: 34294574]

XSL•FO
RenderX

7. Majhail NS, Mau LW, Denzen EM, Arneson TJ. Costs of autologous and allogeneic hematopoietic cell transplantation in the United States: a study using a large national private claims database. Bone Marrow Transplant 2013 Feb;48(2):294-300 [FREE Full text] [doi: 10.1038/bmt.2012.133] [Medline: 22773126]

8. Dunavin N, Mau L, Meyer CL, Divine C, Abdallah A, Leppke S, et al. Health care reimbursement, service utilization, and outcomes among Medicare beneficiaries with multiple myeloma receiving autologous hematopoietic cell transplantation in inpatient and outpatient settings. Biol Blood Marrow Transplant 2020 Apr;26(4):805-813 [FREE Full text] [doi: 10.1016/j.bbmt.2019.12.772] [Medline: 31917269]

9. Granell M, Senín A, Barata A, Cibeira M, Gironella M, López-Pardo J, Group for the Study of MyelomaAmyloidosis of Catalonia (GEMMAC). Predictors of return to work after autologous stem cell transplantation in patients with multiple myeloma. Bone Marrow Transplant 2021 Dec 17;56(12):2904-2910. [doi: 10.1038/s41409-021-01429-0] [Medline: 34404917]

10. Kent EE, Ambs A, Mitchell SA, Clauser SB, Smith AW, Hays RD. Health-related quality of life in older adult survivors of selected cancers: data from the SEER-MHOS linkage. Cancer 2015 Mar 01;121(5):758-765 [FREE Full text] [doi: 10.1002/cncr.29119] [Medline: 25369293]

11. Wood WA, Deal AM, Abernethy A, Basch E, Battaglini C, Kim YH, et al. Feasibility of frequent patient-reported outcome surveillance in patients undergoing hematopoietic cell transplantation. Biol Blood Marrow Transplant 2013 Mar;19(3):450-459 [FREE Full text] [doi: 10.1016/j.bbmt.2012.11.014] [Medline: 23253558]

12. Frödin U, Börjeson S, Lyth J, Lotfi K. A prospective evaluation of patients' health-related quality of life during auto-SCT: a 3-year follow-up. Bone Marrow Transplant 2011 Oct;46(10):1345-1352. [doi: 10.1038/bmt.2010.304] [Medline: 21113189]

13. Cohen MZ, Rozmus CL, Mendoza TR, Padhye NS, Neumann J, Gning I, et al. Symptoms and quality of life in diverse patients undergoing hematopoietic stem cell transplantation. J Pain Symptom Manage 2012 Aug;44(2):168-180 [FREE Full text] [doi: 10.1016/j.jpainsymman.2011.08.011] [Medline: 22699091]

14. El-Jawahri A, LeBlanc T, VanDusen H, Traeger L, Greer JA, Pirl WF, et al. Effect of inpatient palliative care on quality of life 2 weeks after hematopoietic stem cell transplantation: a randomized clinical trial. JAMA 2016 Nov 22;316(20):2094-2103 [FREE Full text] [doi: 10.1001/jama.2016.16786] [Medline: 27893130]

15. McQuellon R, Duckworth K, Campbell C, Russell G, Miskewicz K, Alsobrooks A. Fear of cancer recurrence, distress, depressive symptoms, and quality of life in hematopoietic stem cell transplantation patients. J Psychosoc Oncol Res Pract 2019;1(2):e12 [FREE Full text] [doi: 10.1097/OR9.0000000000000012]

16. Asrar MM, Lad DP, Bansal D, Prinja S, Khadwal A, Prakash G, et al. Health-related quality of life in transplant eligible multiple myeloma patients with or without early ASCT in the real-world setting. Leuk Lymphoma 2021 Dec;62(13):3271-3277. [doi: 10.1080/10428194.2021.1953011] [Medline: 34269156]

17. Lee SJ, Loberiza FR, Antin JH, Kirkpatrick T, Prokop L, Alyea EP, et al. Routine screening for psychosocial distress following hematopoietic stem cell transplantation. Bone Marrow Transplant 2005 Jan 25;35(1):77-83. [doi: 10.1038/sj.bmt.1704709] [Medline: 15502851]

18. El-Jawahri AR, Vandusen HB, Traeger LN, Fishbein JN, Keenan T, Gallagher ER, et al. Quality of life and mood predict posttraumatic stress disorder after hematopoietic stem cell transplantation. Cancer 2016 Mar 01;122(5):806-812 [FREE Full text] [doi: 10.1002/cncr.29818] [Medline: 26650840]

19. Griffith S, Fenech AL, Nelson A, Greer JA, Temel JS, El-Jawahri A. Post-traumatic stress symptoms in hematopoietic stem cell transplant (HCT) recipients. JCO 2020 May 20;38(15_suppl):7505-7505. [doi: 10.1200/jco.2020.38.15_suppl.7505]

20. Banerjee R, Lazar AA, Dunn L, Knoche J, Lo M, Arora S, et al. Benzodiazepine and zolpidem prescriptions during autologous stem cell transplantation. eJHaem 2021 Feb 06;2(2):276-279. [doi: 10.1002/jha2.148]

21. Mishkin AD, Shapiro PA, Reshef R, Lopez-Pintado S, Mapara MY. Standardized semi-structured psychosocial evaluation before hematopoietic stem cell transplantation predicts patient adherence to post-transplant regimen. Biol Blood Marrow Transplant 2019 Nov;25(11):2222-2227 [FREE Full text] [doi: 10.1016/j.bbmt.2019.06.019] [Medline: 31247314]

22. VanDusen H, Traeger L, Fishbein JN, Keenan T, Greer JA, Pirl WF, et al. Psychological distress during hospitalization for hematopoietic stem cell transplantation to predict lower quality of life and high post-traumatic stress disorder symptoms at 6 months post-transplant. JCO 2015 May 20;33(15_suppl):9557-9557. [doi: 10.1200/jco.2015.33.15_suppl.9557]

23. Johnson PC, Bhatt S, Reynolds MJ, Dhawale TM, Ufere N, Jagielo AD, et al. Association between baseline patient-reported outcomes and complications of hematopoietic stem cell transplantation. Transplant Cell Ther 2021 Jun;27(6):496.e1-496.e5. [doi: 10.1016/j.jtct.2021.02.029] [Medline: 33789836]

24. El-Jawahri A, Traeger L, Greer JA, VanDusen H, Fishman SR, LeBlanc TW, et al. Effect of inpatient palliative care during hematopoietic stem-cell transplant on psychological distress 6 months after transplant: results of a randomized clinical trial. J Clin Oncol 2017 Nov 10;35(32):3714-3721 [FREE Full text] [doi: 10.1200/JCO.2017.73.2800] [Medline: 28926288]

25. Jarden M, Nelausen K, Hovgaard D, Boesen E, Adamsen L. The effect of a multimodal intervention on treatment-related symptoms in patients undergoing hematopoietic stem cell transplantation: a randomized controlled trial. J Pain Symptom Manage 2009 Aug;38(2):174-190 [FREE Full text] [doi: 10.1016/j.jpainsymman.2008.09.005] [Medline: 19345060]

26. Deng G, Giralt S, Chung DJ, Landau H, Siman J, Search B, et al. Acupuncture for reduction of symptom burden in multiple myeloma patients undergoing autologous hematopoietic stem cell transplantation: a randomized sham-controlled trial. Support Care Cancer 2018 Feb;26(2):657-665 [FREE Full text] [doi: 10.1007/s00520-017-3881-7] [Medline: 28920142]

XSL•FO
RenderX

27. Bates D, Bolwell B, Majhail NS, Rybicki L, Yurch M, Abounader D, et al. Music therapy for symptom management after autologous stem cell transplantation: results from a randomized study. Biol Blood Marrow Transplant 2017 Sep;23(9):1567-1572 [FREE Full text] [doi: 10.1016/j.bbmt.2017.05.015] [Medline: 28533058]

28. Cassileth BR, Vickers AJ, Magill LA. Music therapy for mood disturbance during hospitalization for autologous stem cell transplantation: a randomized controlled trial. Cancer 2003 Dec 15;98(12):2723-2729 [FREE Full text] [doi: 10.1002/cncr.11842] [Medline: 14669295]

29. Valdimarsdottir HB, Figueiro MG, Holden W, Lutgendorf S, Wu LM, Ancoli-Israel S, et al. Programmed environmental illumination during autologous stem cell transplantation hospitalization for the treatment of multiple myeloma reduces severity of depression: a preliminary randomized controlled trial. Cancer Med 2018 Sep;7(9):4345-4353 [FREE Full text] [doi: 10.1002/cam4.1690] [Medline: 30099857]

30. Wagland R, Fenlon D, Tarrant R, Howard-Jones G, Richardson A. Rebuilding self-confidence after cancer: a feasibility study of life-coaching. Support Care Cancer 2015 Mar;23(3):651-659. [doi: 10.1007/s00520-014-2399-5] [Medline: 25160492]

31. Yun YH, Kim YA, Lee MK, Sim JA, Nam B, Kim S, et al. A randomized controlled trial of physical activity, dietary habit, and distress management with the Leadership and Coaching for Health (LEACH) program for disease-free cancer survivors. BMC Cancer 2017 May 02;17(1):298 [FREE Full text] [doi: 10.1186/s12885-017-3290-9] [Medline: 28464804]

32. Gell NM, Grover KW, Humble M, Sexton M, Dittus K. Efficacy, feasibility, and acceptability of a novel technology-based intervention to support physical activity in cancer survivors. Support Care Cancer 2017 Apr;25(4):1291-1300. [doi: 10.1007/s00520-016-3523-5] [Medline: 27957621]

33. Srivastava U, Burton S, Lewis M, Patel D, Rasulnia M. Assessing the impact of a digital coaching program for patients with cancer. JCO 2018 Oct 20;36(30_suppl):274-274. [doi: 10.1200/jco.2018.36.30_suppl.274]

34. Smith SK, Kuhn E, O'Donnell J, Koontz BF, Nelson N, Molloy K, et al. Cancer distress coach: pilot study of a mobile app for managing posttraumatic stress. Psychooncology 2018 Jan;27(1):350-353. [doi: 10.1002/pon.4363] [Medline: 28032921]

35. Post L, Liefbroer AI. Reducing distress in cancer patients-a preliminary evaluation of short-term coaching by expert volunteers. Psychooncology 2019 Aug 18;28(8):1762-1766 [FREE Full text] [doi: 10.1002/pon.5111] [Medline: 31107998]

36. Gomersall SR, Skinner TL, Winkler E, Healy GN, Eakin E, Fjeldsoe B. Feasibility, acceptability and efficacy of a text message-enhanced clinical exercise rehabilitation intervention for increasing 'whole-of-day' activity in people living with and beyond cancer. BMC Public Health 2019 Jun 03;19(Suppl 2):542 [FREE Full text] [doi: 10.1186/s12889-019-6767-4] [Medline: 31159752]

37. Gell NM, Grover KW, Savard L, Dittus K. Outcomes of a text message, Fitbit, and coaching intervention on physical activity maintenance among cancer survivors: a randomized control pilot trial. J Cancer Surviv 2020 Feb;14(1):80-88. [doi: 10.1007/s11764-019-00831-4] [Medline: 31776849]

38. Im E, Kim S, Yang YL, Chee W. The efficacy of a technology-based information and coaching/support program on pain and symptoms in Asian American survivors of breast cancer. Cancer 2020 Feb 01;126(3):670-680 [FREE Full text] [doi: 10.1002/cncr.32579] [Medline: 31714598]

39. Lo J, Ballurkar K, Fox S, Tynan K, Luu N, Boyer M, et al. A digital coaching intervention for cancer survivors with job loss: retrospective study. JMIR Cancer 2021 Nov 23;7(4):e31966 [FREE Full text] [doi: 10.2196/31966] [Medline: 34710853]

40. Chaar D, Shin JY, Mazzoli A, Vue R, Kedroske J, Chappell G, et al. A mobile health app (Roadmap 2.0) for patients undergoing hematopoietic stem cell transplant: qualitative study on family caregivers' perspectives and design considerations. JMIR Mhealth Uhealth 2019 Oct 24;7(10):e15775 [FREE Full text] [doi: 10.2196/15775] [Medline: 31651402]

41. Chen C, Mao A, Murray J, Oran B, Qazilbash M, Shah N. Feasibility of a smartphone-based health coaching intervention for patient self-management of nutrition in the post-chemotherapy setting. In: Blood | American Society of Hematology. 2016 Presented at: ASH Annual Meeting; December 2, 2016; San Diego, CA p. 1 URL: https://ashpublications.org/blood/article/128/22/3554/97437/Feasibility-of-a-Smartphone-Based-Health-Coaching [doi: 10.1182/blood.v128.22.3554.3554]

42. Amonoo HL, Kurukulasuriya C, Chilson K, Onstad L, Huffman JC, Lee SJ. Improving quality of life in hematopoietic stem cell transplantation survivors through a positive psychology intervention. Biol Blood Marrow Transplant 2020 Jun;26(6):1144-1153 [FREE Full text] [doi: 10.1016/j.bbmt.2020.02.013] [Medline: 32088365]

43. Chen YR, Schulz PJ. The effect of information communication technology interventions on reducing social isolation in the elderly: a systematic review. J Med Internet Res 2016 Jan 28;18(1):e18 [FREE Full text] [doi: 10.2196/jmir.4596] [Medline: 26822073]

44. For Life Sciences | A Better Way to Better Health. Pack Health. URL: https://packhealth.com/life-sciences/ [accessed 2022-02-16]

45. National Board for Health and Wellness Coaching. NBHWC. URL: https://nbhwc.org/ [accessed 2022-02-17]

46. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]

47. Ruggiero L. Transtheoretical model: applications in the prevention and treatment of cancer. Med Pediatr Oncol 1998;Suppl 1:69-74. [doi: 10.1002/(sici)1096-911x(1998)30:1+<69::aid-mpo10>3.0.co;2-w] [Medline: 9659949]

48. Lee HY, Lee MH, Sharratt M, Lee S, Blaes A. Development of a mobile health intervention to promote Papanicolaou tests and human papillomavirus vaccination in an underserved immigrant population: a culturally targeted and individually

tailored text messaging approach. JMIR Mhealth Uhealth 2019 Jun 06;7(6):e13256 [FREE Full text] [doi: 10.2196/13256] [Medline: 31199340]

49. Cheek C, Piercy KW, Grainger S. Leaving home: how older adults prepare for intensive volunteering. J Appl Gerontol 2015 Mar;34(2):181-198. [doi: 10.1177/0733464813480268] [Medline: 24652874]

50. Hagan TL, Arida JA, Hughes SH, Donovan HS. Creating individualized symptom management goals and strategies for cancer-related fatigue for patients with recurrent ovarian cancer. Cancer Nurs 2017;40(4):305-313 [FREE Full text] [doi: 10.1097/NCC.0000000000000407] [Medline: 27333127]

51. Holt KA, Mogensen O, Jensen PT, Hansen DG. Goal setting in cancer rehabilitation and relation to quality of life among women with gynaecological cancer. Acta Oncol 2015 Nov;54(10):1814-1823. [doi: 10.3109/0284186X.2015.1037009] [Medline: 25943136]

52. Chan K, Wong FKY, Lee PH. A brief hope intervention to increase hope level and improve well-being in rehabilitating cancer patients: a feasibility test. SAGE Open Nurs 2019;5:2377960819844381 [FREE Full text] [doi: 10.1177/2377960819844381] [Medline: 33415238]

53. Steffen LE, Cheavens JS, Vowles KE, Gabbard J, Nguyen H, Gan GN, et al. Hope-related goal cognitions and daily experiences of fatigue, pain, and functional concern among lung cancer patients. Support Care Cancer 2020 Feb 1;28(2):827-835 [FREE Full text] [doi: 10.1007/s00520-019-04878-y] [Medline: 31152302]

54. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]

55. Jensen RE, Moinpour CM, Potosky AL, Lobo T, Hahn EA, Hays RD, et al. Responsiveness of 8 Patient-Reported Outcomes Measurement Information System (PROMIS) measures in a large, community-based cancer study cohort. Cancer 2017 Jan 01;123(2):327-335 [FREE Full text] [doi: 10.1002/cncr.30354] [Medline: 27696377]

56. Yost KJ, Eton DT, Garcia SF, Cella D. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. J Clin Epidemiol 2011 May;64(5):507-516 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.11.018] [Medline: 21447427]

57. Bryant AL, Deal AM, Battaglini CL, Phillips B, Pergolotti M, Coffman E, et al. The effects of exercise on patient-reported outcomes and performance-based physical function in adults with acute leukemia undergoing induction therapy: Exercise and Quality of Life in Acute Leukemia (EQUAL). Integr Cancer Ther 2018 Jun 24;17(2):263-270 [FREE Full text] [doi: 10.1177/1534735417699881] [Medline: 28627275]

58. Riba MB, Donovan KA, Andersen B, Braun I, Breitbart WS, Brewer BW, et al. Distress Management, Version 3.2019, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw 2019 Oct 01;17(10):1229-1249 [FREE Full text] [doi: 10.6004/jnccn.2019.0048] [Medline: 31590149]

59. Runaas L, Hoodin F, Munaco A, Fauer A, Sankaran R, Churay T, et al. Novel health information technology tool use by adult patients undergoing allogeneic hematopoietic cell transplantation: longitudinal quantitative and qualitative patient-reported outcomes. JCO Clin Cancer Inform 2018 Dec;2(2):1-12 [FREE Full text] [doi: 10.1200/CCI.17.00110] [Medline: 30652535]

60. Racioppi A, Dalton T, Ramalingam S, Romero K, Ren Y, Bohannon L, et al. Assessing the feasibility of a novel mHealth app in hematopoietic stem cell transplant patients. Transplant Cell Ther 2021 Feb;27(2):181.e1-181.e9. [doi: 10.1016/j.jtct.2020.10.017] [Medline: 33830035]

61. Marthick M, Janssen A, Cheema BS, Alison J, Shaw T, Dhillon H. Feasibility of an interactive patient portal for monitoring physical activity, remote symptom reporting, and patient education in oncology: qualitative study. JMIR Cancer 2019 Nov 28;5(2):e15539 [FREE Full text] [doi: 10.2196/15539] [Medline: 31778123]

62. Saleem M, Kühne L, De Santis KK, Christianson L, Brand T, Busse H. Understanding engagement strategies in digital interventions for mental health promotion: scoping review. JMIR Ment Health 2021 Dec 20;8(12):e30000 [FREE Full text] [doi: 10.2196/30000] [Medline: 34931995]

63. Carlo AD, Hosseini Ghomi R, Renn BN, Strong MA, Areán PA. Assessment of real-world use of behavioral health mobile applications by a novel stickiness metric. JAMA Netw Open 2020 Aug 03;3(8):e2011978 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.11978] [Medline: 32744628]

64. Hershman DL, Unger JM, Hillyer GC, Moseley A, Arnold KB, Dakhil SR, et al. Randomized trial of text messaging to reduce early discontinuation of adjuvant aromatase inhibitor therapy in women with early-stage breast cancer: SWOG S1105. J Clin Oncol 2020 Jul 01;38(19):2122-2129 [FREE Full text] [doi: 10.1200/JCO.19.02699] [Medline: 32369401]

65. Costa LJ, Huang J, Hari PN. Disparities in utilization of autologous hematopoietic cell transplantation for treatment of multiple myeloma. Biol Blood Marrow Transplant 2015 Apr;21(4):701-706 [FREE Full text] [doi: 10.1016/j.bbmt.2014.12.024] [Medline: 25555447]

66. Jayakrishnan TT, Bakalov V, Chahine Z, Lister J, Wegner RE, Sadashiv S. Disparities in the enrollment to systemic therapy and survival for patients with multiple myeloma. Hematol Oncol Stem Cell Ther 2021 Sep;14(3):218-230 [FREE Full text] [doi: 10.1016/j.hemonc.2020.09.005] [Medline: 33069693]

67. Duckworth KE, Forti AM, Russell GB, Naik S, Hurd DD, McQuellon RP. Hematopoietic stem cell transplant candidate and designated proxy distress levels prior to hematopoietic stem cell transplantation. Am J Hosp Palliat Care 2014 Dec;31(8):853-856 [FREE Full text] [doi: 10.1177/1049909113508217] [Medline: 24142596]

68.   Ransom S, Jacobsen PB, Booth-Jones M. Validation of the Distress Thermometer with bone marrow transplant patients. Psychooncology 2006 Jul;15(7):604-612. [doi: 10.1002/pon.993] [Medline: 16208733]

69.   Baliousis M, Rennoldson M, Snowden JA. Psychological interventions for distress in adults undergoing haematopoietic stem cell transplantation: a systematic review with meta-analysis. Psychooncology 2016 Apr;25(4):400-411. [doi: 10.1002/pon.3925] [Medline: 26249330]

70.   Mitchell M, White L, Oh P, Alter D, Leahey T, Kwan M, et al. Uptake of an incentive-based mHealth app: process evaluation of the Carrot Rewards App. JMIR Mhealth Uhealth 2017 May 30;5(5):e70 [FREE Full text] [doi: 10.2196/mhealth.7323] [Medline: 28559224]

71.   Shaw BE, Lee SJ, Horowitz MM, Wood WA, Rizzo JD, Flynn KE. Can we agree on patient-reported outcome measures for assessing hematopoietic cell transplantation patients? A study from the CIBMTR and BMT CTN. Bone Marrow Transplant 2016 Sep;51(9):1173-1179. [doi: 10.1038/bmt.2016.113] [Medline: 27159181]

72.   Sidana S, Thanarajasingam G, Griffin J, Thompson C, Burtis M, Warsame R. Patient experience of chimeric antigen receptor (CAR)-T cell therapy vs stem cell transplant: longitudinal patient reported adverse events, cognition and quality of life. In: Blood | American Society of Hematology. 2019 Presented at: ASH Annual Meeting; December 13, 2019; Orlando, FL p. 794. [doi: 10.1182/blood-2019-121715]

Abbreviations

DLC: digital life coaching
ePRO: Electronic patient-reported outcomes
MM: multiple myeloma
NCCN: National Comprehensive Cancer Network
PROMIS: Patient-reported Outcome Measurement Information System
QOL: quality of life
REDCap: Research Electronic Data Capture
SCT: stem cell transplantation
SMART: Specific, Measurable, Achievable, Relevant, and Time-based

<u>Original Paper</u>

Token Economy–Based Hospital Bed Allocation to Mitigate Information Asymmetry: Proof-of-Concept Study Through Simulation Implementation

Shusuke Hiragi[1,2,3], MD, MBA, PhD; Jun Hatanaka[2], MSci; Osamu Sugiyama[4], PhD; Kenichi Saito[1], MD; Masayuki Nambu[5], PhD; Tomohiro Kuroda[1,2], PhD

[1]Division of Medical Informatics and Administration Planning, Kyoto University Hospital, Kyoto, Japan

[2]Graduate School of Informatics, Kyoto University, Kyoto, Japan

[3]Division of Health Science, Tazuke Kofukai Medical Research Institute, Osaka, Japan

[4]Department of Real World Data Research and Development, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[5]Preemptive Medicine & Lifestyle-Related Disease Research Center, Kyoto University Hospital, Kyoto, Japan

Corresponding Author:
Shusuke Hiragi, MD, MBA, PhD
Division of Health Science
Tazuke Kofukai Medical Research Institute
2-4-10 Ohgimachi Kita-ku
Osaka, 5308480
Japan
Phone: 81 6 6312 1221
Email: shiragi@kuhp.kyoto-u.ac.jp

Abstract

Background: Hospital bed management is an important resource allocation task in hospital management, but currently, it is a challenging task. However, acquiring an optimal solution is also difficult because intraorganizational information asymmetry exists. Signaling, as defined in the fields of economics, can be used to mitigate this problem.

Objective: We aimed to develop an assignment process that is based on a token economy as signaling intermediary.

Methods: We implemented a game-like simulation, representing token economy–based bed assignments, in which 3 players act as ward managers of 3 inpatient wards (1 each). As a preliminary evaluation, we recruited 9 nurse managers to play and then participate in a survey about qualitative perceptions for current and proposed methods (7-point Likert scale). We also asked them about preferred rewards for collected tokens. In addition, we quantitatively recorded participant pricing behavior.

Results: Participants scored the token economy–method positively in staff satisfaction (3.89 points vs 2.67 points) and patient safety (4.38 points vs 3.50 points) compared to the current method, but they scored the proposed method negatively for managerial rivalry, staff employee development, and benefit for patients. The majority of participants (7 out of 9) listed human resources as the preferred reward for tokens. There were slight associations between workload information and pricing.

Conclusions: Survey results indicate that the proposed method can improve staff satisfaction and patient safety by increasing the decision-making autonomy of staff but may also increase managerial rivalry, as expected from existing criticism for decentralized decision-making. Participant behavior indicated that token-based pricing can act as a signaling intermediary. Given responses related to rewards, a token system that is designed to incorporate human resource allocation is a promising method. Based on aforementioned discussion, we concluded that a token economy–based bed allocation system has the potential to be an optimal method by mitigating information asymmetry.

XSL•FO
RenderX

Introduction

Hospital bed management is an important resource allocation task for better patient care and sustainable hospital management [1]. In particular, efficient bed utilization is a key driver for hospital revenue [2] and health care system management [3,4]. The management of hospital finances has become difficult in Japan [5,6] and worldwide [7-9]; thus, to cope, effective resource management is essential.

Poor bed management wastes time, resulting in longer wait times for patients, the reduction of employee (especially ward nurses) satisfaction [10], and adverse effects on patient safety [11,12].

The current bed control process begins when physicians admit patients to the hospital and asked bed allocation managers to find adequate beds [13]. Allocation managers search the wards to find the best fit for the patient. Then, either the *negotiation-based method*, wherein allocation managers negotiate with frontline ward managers, who sometimes decline the request, to form consensus and make a final decision [14], or the *command-based method*, wherein allocation managers have the authority to force wards to accept the request [15], emerges. The negotiation-based method takes more time [14], while the command-based method hinders employee satisfaction by involuntarily increasing the patient-to-nurse ratio, which has been reported to increase staff dissatisfaction [16]. In fact, these methods are not discrete but form a continuum, as procedures, in reality, both have negotiation-based as well as command-based aspects.

Conflict is defined as the struggle that arises when the goal-directed behavior of a person or group blocks that of another [17]. Both bed allocation methods include a systemic structure that can cause intraorganizational conflict between allocation managers and frontline ward workers. Previous reports mentioned that mediation of information asymmetry is an effective form of conflict management [18], and often, information asymmetry is observed in real practice between allocation managers, who have broad but superficial views of the bed occupancy situation for the whole hospital, and managers of individual wards, who tend to only see their situation but perfectly know their capacity.

Previous studies [19,20] in operational research have tried to solve the bed assignment problem with mathematical modeling, which is based on the assumption that computers can calculate optimal solutions for command-based bed allocation to satisfy any stakeholder requirements; despite this, no real-world solutions are in wide use, because of the implementation difficulties arising from inputting data from a variety of sources and convincing users to accept computed suggestions [19]. This situation indicates that collecting sufficient information is difficult, which can lead to information asymmetry, resulting in intraorganizational conflict.

In the field of economics and management, *signaling*, whereby one party credibly conveys some information about itself to another [21], is a solution for information asymmetry [22]. However, requiring too much information from frontline staff is not feasible in practice; therefore, simple information to be expressed is needed. Herein, we propose the implementation of a token economy–based operation for bed allocation, in which the allocation process is regarded as a virtual currency transaction. Token economy was originally developed for behavior modification in psychology [23] and education [24] and is now also used by businesses [25] that intend to manage user behavior. We hypothesize that this system has the advantages of both negotiation-based and command-based allocation systems.

We developed a token economy–based system for bed allocation and conducted a preliminary evaluation of the system with a game-like simulation followed by a brief survey about its effectiveness.

Methods

Token Economy–Based Bed Assignment System

Overview

The current bed assignment process begins when physicians decide to admit a patient to the hospital and the allocation manager negotiates with ward staff to find an available bed (Figure 1).

However, we propose that the bed allocation process can be considered as transactions of nursing contributions between physicians and nursing staff. Each ward in the hospital is asked to set a *price* (ie, define the value for their nursing service) using a virtual currency (ie, token) in advance. When physicians admit patients, the hospital allocates a certain number of tokens to the physicians as consideration for admission. Then, using *prices* and their preferences, such as ward specialty, physicians decide where patients will be assigned. Once the deal is finalized, the bed allocation process is completed (Figure 1). In this process, allocation managers do not have a role. By repeating this procedure, each stakeholder can accumulate tokens, to be exchanged for something that has value to them, and ward *price* setting acts as a signaling of availability, which removes information asymmetry to result in optimal bed allocation.

Figure 1. Operation flowchart of current (upper) and proposed (lower) bed allocation methods.
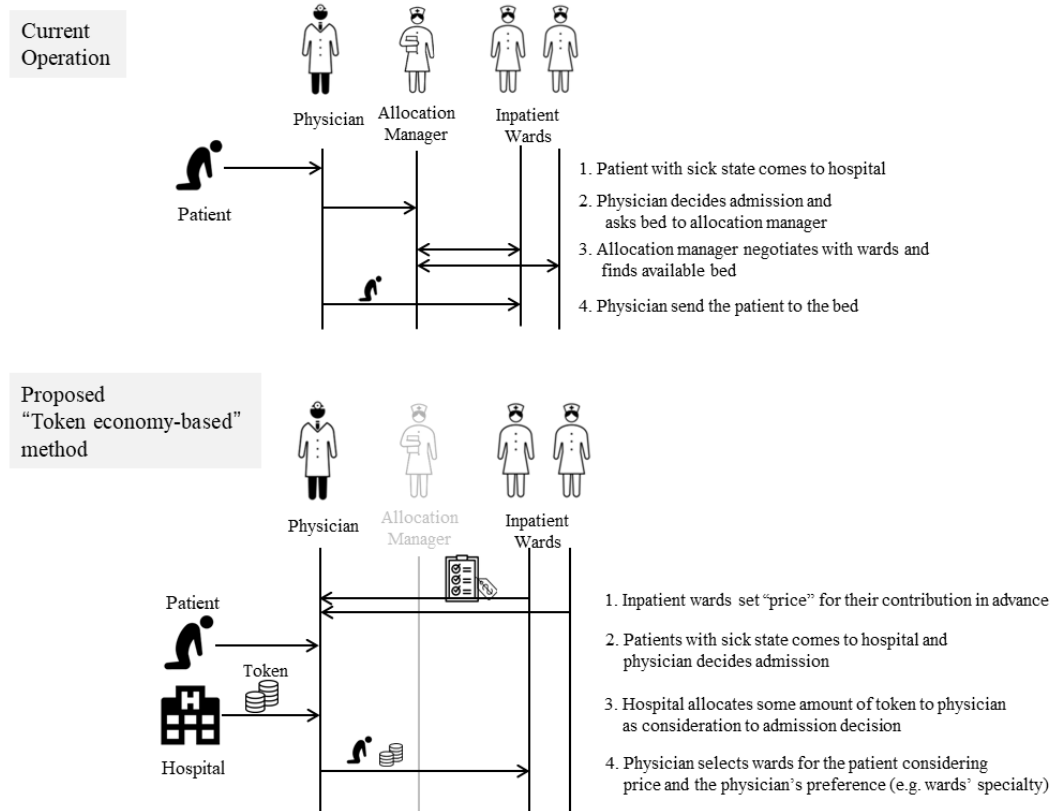
Implementation of Simulation System

To evaluate the token economy–based bed allocation system, we developed a game-like simulation. We developed a web-based simulation for nurse managers (who represent the nursing staff on the wards). The behavior of physicians was emulated by computer agents.

The scenario consisted of a small hospital with 3 wards, each having 40 beds. For simplicity, we used the assumption that there are 3 illnesses (ie, gastric ulcer, pneumonia, and heart failure) with 3 severities (ie, mild, moderate, and severe). To represent the real-world allocation of specialties, we also defined the assumption that each virtual ward was used to manage a single type of illness: ward A, gastric ulcer; ward B, pneumonia; and ward C, heart failure (Figure 2).

Each virtual ward had a *workload* parameter, that reflected nursing staff workloads. This information was only seen by the relevant ward, thus representing information asymmetry. The workload increased in response to bed occupancy, weighted by patient illness and severity: *workload* increased by 1.0, 1.25, and 1.5 points when the ward received a patient with mild, moderate, and severe, respectively, corresponding illness; however, when the patient's illness was not aligned with the ward's specialty, *workload* increased by an additional 50%—1.5, 1.8, and 2.25 points for mild, moderate, and severe illnesses, respectively, not corresponding to that of the ward specialty (Table S1 in Multimedia Appendix 1).

Figure 2. Game scenario: the hospital consisted of 3 wards, with 40 beds each, and patients with 3 diseases with 3 levels of severity.
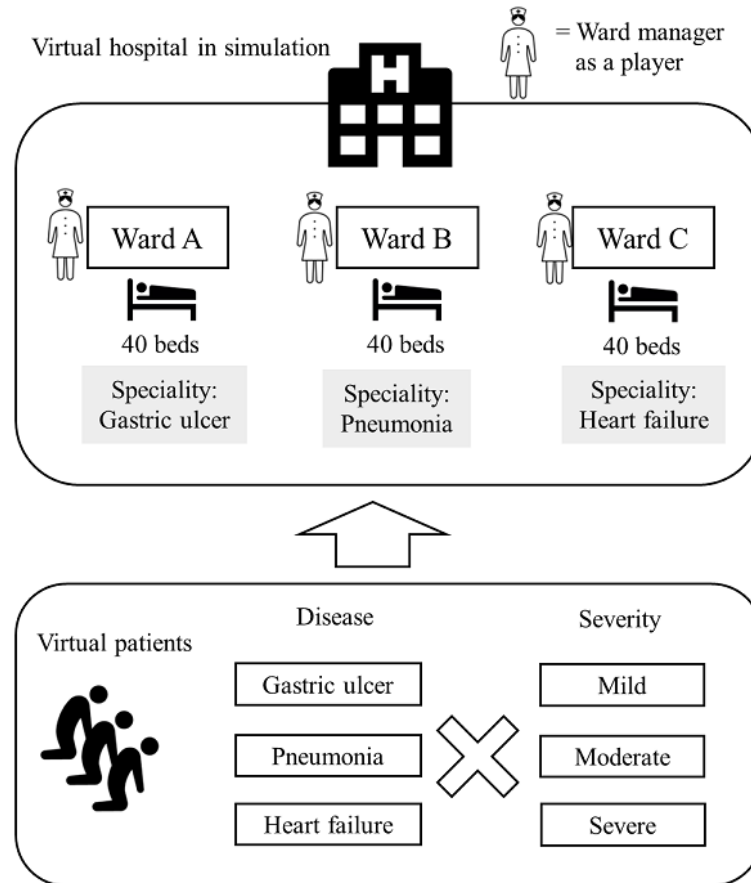
User Interface

The simulation consisted of day-level repetition (Figure 3). At the beginning of a day, nurse managers (players) were asked to set a *price* for each illness and severity. They could see the current bed occupancy and staff workload of their ward, but they could only see the number of patients admitted to other wards (Figure S1 in Multimedia Appendix 1). Players were then asked asked to set *prices* for 9 patterns of patients (Figure S2 in Multimedia Appendix 1). Each day, 7 patients appeared, with random illnesses and severities (Figure S3 in Multimedia Appendix 1) and were allocated to the ward selected by physicians—a computer agent programmed to select the ward that proposed the lowest *price* unless the ward specializing in the patient's illness was available, in which case, a certain additional *price* was acceptable. The additional rate was set as a uniform random number between 1.0 (no additional fee) and 2.0 for every admission decision. Patients were randomly discharged from wards at a rate equivalent to a mean of 14 days, so that vacant beds appeared in wards. We repeated the simulation 10 times, representing 10 real-world days. All simulations were implemented with Python (version 3.4.5) and Django (version 2.0.13).

Figure 3. Flowchart illustrating the game. The gray area indicates 1 day in the simulation, with the game consisted of 10 virtual days.
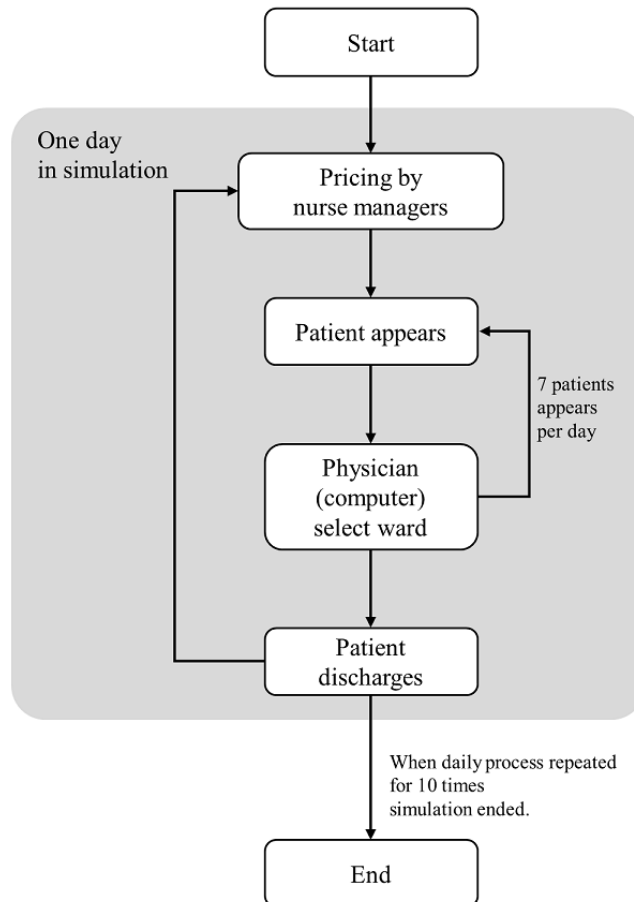
Evaluation

Participants

We recruited 9 nurse managers from a single university-affiliated hospital in Japan. We split them into 3 groups with 3 participants each. After receiving informed consent, we asked the nurse managers to participate in 2 sessions of the game-like simulation. During the first session, participants were told that the player who earned the most tokens would be the winner. For the second session, participants were instructed to assume they can receive the form of compensation they wish and that they are required to make decisions as nurse managers, by considering aspects such as staff satisfaction, patient safety, and equality, as they did in their daily practice.

We conducted 3 surveys with the participants: before, between, and after sessions (Figure 4). Using a similar question format about perceptions relating to manager satisfaction, staff satisfaction, benefit for patients, patient safety, timeliness of decision making, extent of managerial rivalry, extent of managerial control from hospital administrators, effect on revenues, employee development, consistency with

organizational mission (which is equal to the whole hospital's mission), and consistency with the goal of each ward as independent division (Table 1), the before-session survey asked participants about their perceptions of the current bed assignment practices, whereas the after-session survey asked participants about their perceptions of the token economy–based assignment method. Each response used a 7-point Likert scale (1, strongly disagree, to 7, strongly agree). We developed the questionnaire to examine the system's effect on hospital management through resource allocation decision-making, given the advantages and disadvantages that have been identified by previous researchers [26], of decentralized decision-making in an organization. In addition, we considered Balanced scorecard, which is well-known organizational performance assessment tool [27]. In particular, questionnaire about employee development was set to evaluate this method could develop ward staffs' managerial ability. We did not perform statistical testing, due to the limited number of participants.

For the interim survey, we asked participants about desired compensation regarding the token—"Please enumerate what you want to get as a consideration of tokens. List at least two, and at most five, items with giving priority."

Figure 4. Evaluation flowchart.
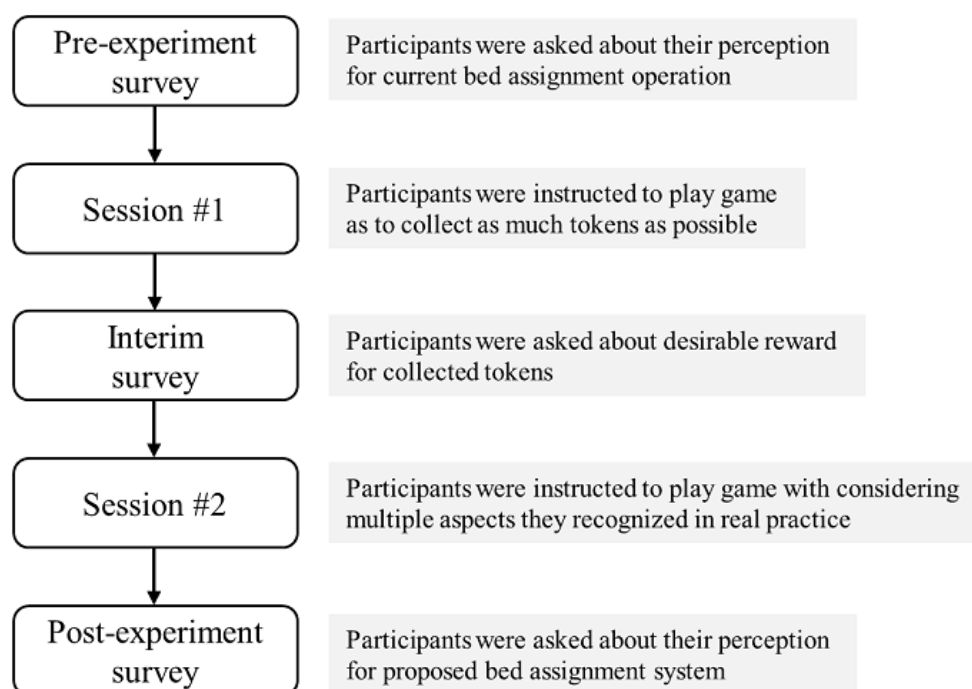
Table 1. Before- and after-session survey questions.

Topic	Current bed management practices	Proposed token economy–based system
Managers' satisfaction	Are you satisfied with it?	Will you be satisfied with it?
Staff satisfaction	Do you think your staff is satisfied with it?	Do you think your staff will be satisfied with it?
Benefit for patients	Do you think it is beneficial for patients?	Do you think it will be beneficial for patients?
Patient safety	Do you think it has problems from the patient safety point of view?	Do you think it will have problems from the patient safety point of view?
Timeliness	Do you think it offers prompt decision-making?	Do you think it will offer prompt decision-making?
Managerial rivalry	Do you think it evokes managerial rivalry?	Do you think it will evoke managerial rivalry?
Control	Do you think managerial governance is maintained though it?	Do you think managerial governance will be maintained though it?
Revenue	Do you think it is optimized for hospital revenue?	Do you think it will be optimized for hospital revenue?
Employee development	Do you think it helps your staffs' employee development?	Do you think it will help your staffs' employee development?
Hospital mission	Do you think it is consistent with the hospital mission?	Do you think it will be consistent with the hospital mission?
Goal of divisions	Do you think it considers the goal of each ward?	Do you think it will consider the goal of each ward?

Analysis of Participants' Price Setting

We recorded *price* list input by participants and *workload* fluctuations. We collected data from 6 sessions (3 groups, each playing 2 sessions). To evaluate this token economy–based method as a signaling mediator, we plotted the associations between *workload* and *price* setting.

Ethics Approval

These experiments were conducted with the approval of the Ethics Committee of Kyoto University Graduate School and Faculty of Medicine (R1972).

Results

Participants

Participants had more than 25 years (mean 31.1 years) of experience as nurses and more than 3 years (mean 11.6 years) as nurse managers at the time of experiment (Table 2).

Table 2. Demographic backgrounds of participants.

Characteristics	Value
Gender (n=9), n (%)	
Male	1 (11)
Female	8 (89)
Experience (years), mean (range)	
Nurse	31.1 (26-34)
Nurse manager	11.6 (4-18)

Survey Results

Staff satisfaction for current operation practices and for the proposed method were at 2.67 and 3.89 points on average, respectively. Conversely, they scored benefit for patients in current practices and for the proposed method at 4.44 and 3.56 points, respectively. Participants gave favorable responses for the proposed method when compared with those for current practices for manager satisfaction, patient safety, timely decision-making, and consistency with goal of divisions; however, they gave unfavorable responses for the proposed method compared with those of current practices for the extent of managerial rivalry, the extent of control from hospital administration, employee development, and consistency with hospital mission (Table 3 and Figure 5).

Most participants (7 out of 9) listed additional human resources as preferred compensation for tokens. Financial compensation, such as salary increases or bonuses, ranked in the top 3 for some participants (4 out of 9). Only 1 participant did not list items related to human resources; this participant listed commendation from the hospital as the top priority (Table 4).

Table 3. Before- (current) and after-session (proposed) survey scores; items were rated on a 7-point Likert scale (1, strongly disagree, to 7, strongly agree).

	Current bed management practices, mean	Proposed token economy–based system, mean	Difference
Manager satisfaction	3.44	3.78	0.33
Staff satisfaction	2.67	3.89	1.22
Benefit for patients	4.44	3.56	−0.89
Patient safety[a]	3.50	4.38	0.88
Timeliness	3.67	4.00	0.33
Managerial rivalry[b]	2.78	3.67	0.89
Control	4.44	3.44	−1.00
Revenue	3.78	3.78	0.00
Employee development	4.11	3.44	−0.67
Hospital mission	4.56	4.11	−0.44
Goal of divisions	3.00	3.67	0.67

[a]There was a missing response; therefore, n=8 for this item.

[b]Only for this item, a higher score means a negative response.

Figure 5. Differences between perceptions in current bed management practices and those for the proposed bed management method. Since a higher score for managerial rivalry indicates a negative response, we inverted the raw data. There was 1 missing response for patient safety.
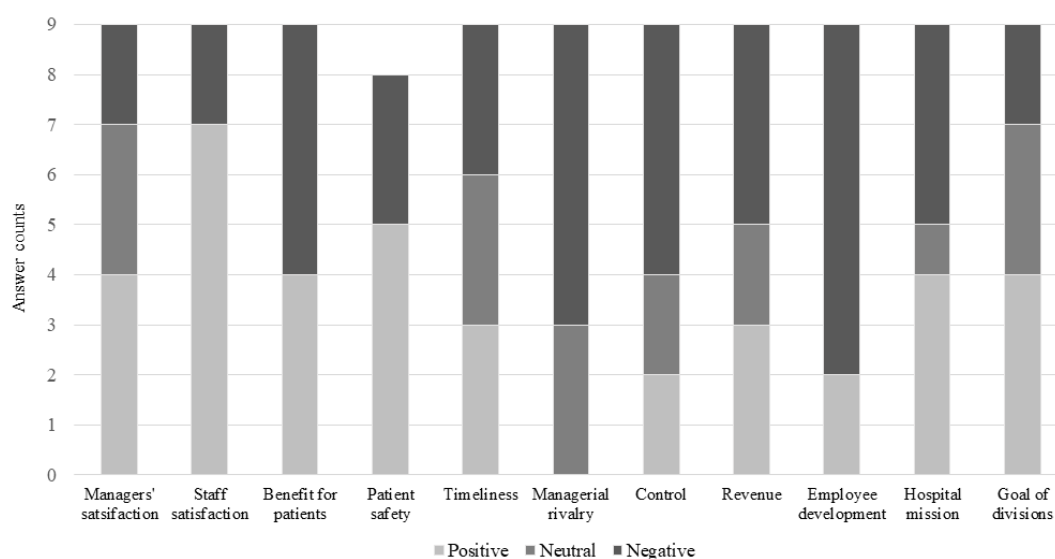
Table 4. Participants' response to the question asking the desirable rewards for collected token.

Participant	Priority				
	First	Second	Third	Fourth	Fifth
1	Human resources (nurses)	Travel fee for workshop	Travel fee for conference	Spare time or bonus	Improved work environment
2	Human resources (nurses)	Travel fee for conference	Improved work environment	Human resources (nurse assistants)	Equipment
3	Human resources (nurses)	Vacation	Bonus	Reduction of overtime	Support for mental condition
4	Human resources (nurses)	Bonus	Office supplies	Equipment	Human resources (doctors)
5	Allowance	Human resources (nurses)	Vacation	Spare time	Recreation fee
6	Human resources (nurse)	Human resources (nurse assistants)	Improved work environment	Education materials	__a
7	Human resources (nurses)	Salary	Human resources (doctors)	Improved work environment	—
8	Commendation	Equipment	Recreation fee	—	—
9	Human resources (nurses)	Spare time	Equipment	—	—

[a]Participants were asked to list between 2 and 5 items; therefore, some entries are blank.

Workloads and Price Setting

When average *workload* increased, participants typically set higher prices for each illness; however, this did not occur every time (Figure 6). Every group, especially Group A, showed a more apparent association between price and workload during the second session than that in first session (Figures S4 and S5 in Multimedia Appendix 1). There was a similar tendency for mild illnesses; however, we could not evaluate quantitative association due to limited data.

Figure 6. Association between price and workload. The upper graphs show the first session, in which players were asked to collect as many tokens as possible, and the lower graphs show the second session, in which players were asked to make decisions as nurse managers in real practice. A darker gray indicates a higher workload.
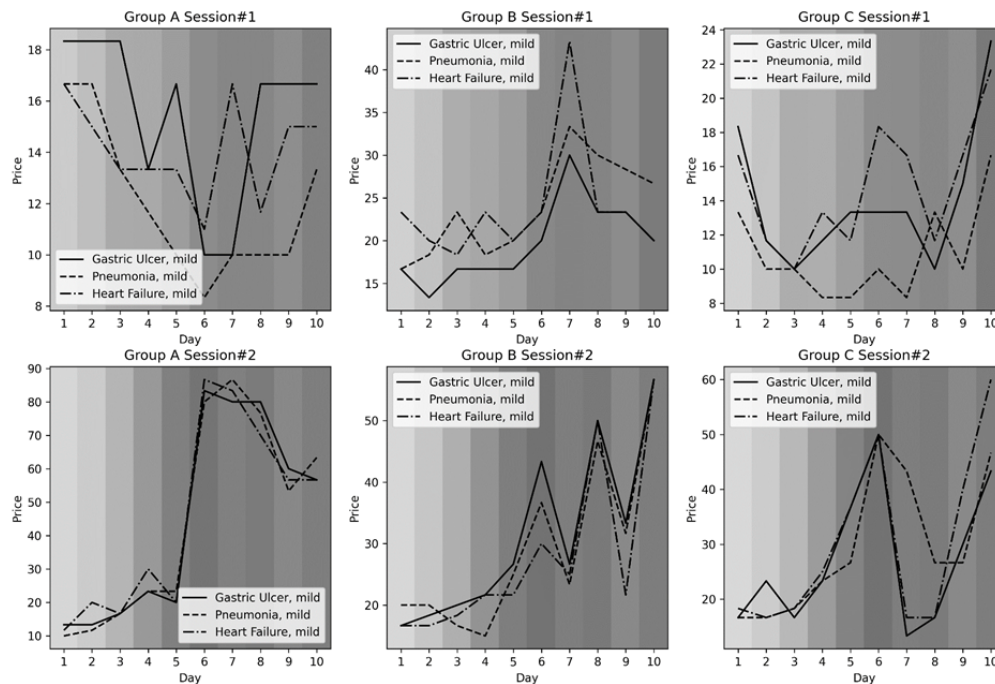
Discussion

General

A preliminarily examination of practicing nurse managers' perception on the feasibility of using a token economy–based bed allocation system to mitigate information asymmetry indicated that the system may enhance staff satisfaction and patient safety but may worsen managerial rivalry and inhibit staff development. There is also potential that hospital administrators can utilize our system as a strategy development tool for human resource allocation.

In mathematical modeling studies of bed assignment processes, patient flow, and bed availability [19,20], there is the implicit assumption that there is no information asymmetry, which allows an optimized solution to be found; however, as those studies [19,20] indicated, communication problems between planners and frontline workers remained. That is, information asymmetry diminished the efficacy of such methods in practice. The impossibility of the planner being able to collect complete information has been widely acknowledged in economics (ie, calculation controversy [28]) and by military strategists (fog of war [29]). Likewise, in real practice at hospitals, there is too much apparent or latent information that can be considered in bed assignment process; thus, solving the bed assignment operation problem with mathematical calculation is fundamentally unfeasible.

To the best of our knowledge, no previous research has focused on information asymmetry in bed allocation or hospital management. We identified optimized bed assignment as the target behavior and designed our token economy with transactions occurring between physicians and impatient wards. Unlike token economy designs, in which psychiatrists or

teachers define target behavior unilaterally, our design made both stakeholder parties represent desirable behaviors with tokens. Even though the sample size was limited, associations between the information that participants had (*workload*) and the information that participants broadcast (*price*) demonstrated that our token economy–based has potential to mitigate information asymmetry in bed allocation management.

Participants positively rated staff satisfaction, patient safety, and consistency with divisional goals. Autonomy is said to be one of the important factors in improving employee satisfaction in hospitals [30], and characteristics of the proposed method—each frontline division can broadcast their status in a 1-dimensional simplified manner, in the form of pricing—may enhance the autonomy of divisions, resulting in positive effects, which may explain our survey results. Employee satisfaction increases organizational loyalty and employee retention [31], which is critical for human resource management generally and specifically in the nursing filed [32,33]. In addition, patient safety improves when employees are satisfied with their workplace [34] and vice versa [35].

Participants negatively rated the effect on the intensity of managerial rivalry, employee development, and the benefits for patients. Managerial rivalry is said to be a disadvantage of decentralized decision-making;. However, a decentralized method is also said to have positive effects on employee development, by allowing autonomous decision-making [26].

Participants were nurse managers in a teaching hospital, where many employees have limited experience as nurses to make adequate decisions independently; therefore, participants, as managers, might be afraid that their staff would make selfish choices [36]. Given this, prior to being implementing this token economy–based systems in practice, staff decision-making

capabilities should be evaluated. Regarding benefit for patients, our method did not take ward specialty into consideration when selecting admission for the sake of simplicity; therefore, participants thought patients had higher possibilities of being assigned to nonspecialized wards than the current method. To address this problem, rule-based restrictions can be considered in future system designs.

When asked about desirable rewards for tokens, most participants indicated human resources (in particular, additional nursing staff). Rewards for tokens are called "back-up reinforcer(s)" [37] when used in behavior modification fields, and are key factors in the method's design [38]. Hospitals in Japan are suffering from nurse shortages [39], and efficient human resource allocation is needed. Wards (or divisions) that collect more tokens can be regarded as creating more value for the hospital. Therefore, allocating more resources to those wards is reasonable, especially given that hospitals are a labor-intensive industry [40], in nurses play central role [41]. Through such implementation design, a token economy–based system may be able to optimize resource allocation and improve staff motivation.

Although we chose bed allocation as the target problem, hospitals have many other resource allocation problems to be solved. The proposed token economy–based method may be applied to other resource allocation problems in hospitals in the future, with appropriate token economy design.

Limitations

The study had some limitations because this research showed only preliminary implications for efficiency in applying a token economy–based system to bed assignment processes. First, the assumptions used in our simulation were too simple compared with those in real practice. That is, we only considered emergency admissions, whereas in reality, hospitals also receive elective admissions. The proposed token economy–based system may cause confusion for nurse managers whether emergency and elective cases should be prioritized at first, but they will be able to adjust pricing behavior by referring elective admission lists. We can, theoretically, expect that the adjustment may result in convergence to adequate allocation via market mechanism. Second, participants were recruited from a single university-affiliated teaching hospital, thus the perceptions of the participants may be different from the perceptions of staff in general hospital fields. Third, we only recruited 9 participants, due to limited resources, and the limited sample size may result in impaired external validity.

Based on these limitations, immediate implementation of our method into real practice is probably unfeasible. However, by showing the constraints of previous studies [19,20] and proposing an innovative method, we make an important contribution.

Conclusion

A token economy–based bed allocation system has the potential to be an effective method to innovatively solve bed allocation problems in hospitals, by mitigating intraorganizational information asymmetry, and improve staff satisfaction, by allowing the autonomy of frontline professionals.

Authors' Contributions

SH provided the research idea, wrote the manuscript, and led the preparation of figures and tables. JH designed and conducted experiments. OS supervised programming of the simulation game. KS helped with data analysis and preparation of the manuscript, figures, and tables. MN and TK led the discussion and the team.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables and figures.
[PDF File (Adobe PDF File), 1314 KB - formative_v6i3e28877_app1.pdf]

References

1. Schäfer F, Walther M, Hübner A. Patient-bed allocation in large hospitals. In: Proceedings of the Third International Conference on Health Care Systems Engineering. 2017 Presented at: International Conference on Health Care Systems Engineering; May 29-31; Florence, Italy p. 299-300. [doi: 10.1007/978-3-319-66146-9_28]

2. Forneas N. Improving hospital bed management with AI. IBM Customer Stories. 2018. URL: https://www.ibm.com/blogs/client-voices/improving-hospital-bed-management-ai/ [accessed 2019-03-26]

3. Phillip PJ, Mullner R, Andes S. Toward a better understanding of hospital occupancy rates. Health Care Financ Rev 1984;5(4):53-61 [FREE Full text] [Medline: 10310946]

4. Belciug S, Gorunescu F. Improving hospital bed occupancy and resource utilization through queuing modeling and evolutionary computation. J Biomed Inform 2015 Feb;53:261-269 [FREE Full text] [doi: 10.1016/j.jbi.2014.11.010] [Medline: 25433363]

5. Zhang X, Oyama T. Investigating the health care delivery system in Japan and reviewing the local public hospital reform. Risk Manag Healthc Policy 2016;9:21-32 [FREE Full text] [doi: 10.2147/RMHP.S93285] [Medline: 27051323]

6. Kawaguchi H, Tone K, Tsutsui M. Estimation of the efficiency of Japanese hospitals using a dynamic and network data envelopment analysis model. Health Care Manag Sci 2014 Jun;17(2):101-112 [FREE Full text] [doi: 10.1007/s10729-013-9248-9] [Medline: 23942762]

7. Saleh S, Kaissi A, Semaan A, Natafgi NM. Strategic planning processes and financial performance among hospitals in Lebanon. Int J Health Plann Manage 2013;28(1):e34-e45. [doi: 10.1002/hpm.2128] [Medline: 22859363]

8. Ramamonjiarivelo Z, Weech-Maldonado R, Hearld L, Menachemi N, Epané JP, O'Connor S. Public hospitals in financial distress: is privatization a strategic choice? Health Care Manage Rev 2015;40(4):337-347. [doi: 10.1097/HMR.0000000000000032] [Medline: 25029510]

9. The evolution of smart health care. Deloitte Touche Tohmatsu Limited. 2018. URL: https://www2.deloitte.com/content/dam/Deloitte/tr/Documents/life-sciences-health-care/gx-lshc-hc-outlook-2018.pdf [accessed 2019-03-26]

10. Yamaguchi K, Ikeda S, Muto M, Narita T. Using Data mining to investigate the factors influencing nurses to quit their jobs? from the perspectives of bed management, working environment and job satisfaction? J Japan Assoc Health Care Admin 2013;7(1):55-65. [doi: 10.11202/jaha.7.55]

11. Thomas Schneider AJ, Luuk Besselink P, Zonderland ME, Boucherie RJ, van den Hout WB, Kievit J, et al. Allocating emergency beds improves the emergency admission flow. Interfaces 2018 Aug;48(4):384-394. [doi: 10.1287/inte.2018.0951]

12. McMurdo MET, Witham MD. Unnecessary ward moves. Age Ageing 2013 Sep 28;42(5):555-556. [doi: 10.1093/ageing/aft079] [Medline: 23892919]

13. Kang H, Nembhard H, Rafferty C, DeFlitch C. Patient flow in the emergency department: a classification and analysis of admission process policies. Ann Emerg Med 2014 Oct;64(4):335-342.e8 [FREE Full text] [doi: 10.1016/j.annemergmed.2014.04.011] [Medline: 24875896]

14. Shen Y, Lee LH. Improving the wait time to admission by reducing bed rejections. BMJ Open Qual 2019;8(3):e000710 [FREE Full text] [doi: 10.1136/bmjoq-2019-000710] [Medline: 31414061]

15. Novati R, Papalia R, Peano L, Gorraz A, Artuso L, Canta MG, et al. Effectiveness of an hospital bed management model: results of four years of follow-up. Ann Ig 2017;29(3):189-196 [FREE Full text] [doi: 10.7416/ai.2017.2146] [Medline: 28383610]

16. Aiken LH, Clarke SP, Sloane DM, Sochalski J, Silber JH. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. JAMA 2002;288(16):1987-1993. [Medline: 12387650]

17. George J, Jones G. Power, politics, conflict, and negotiation. In: Understanding and Managing Organizational Behavior. Hoboken, New Jersey: Prentice Hall; Dec 31, 2010:370-399.

18. Rauchhaus RW. Asymmetric information, mediation, and conflict management. World Pol 2011 Jun 13;58(2):207-241. [doi: 10.1353/wp.2006.0027]

19. Thomas BG, Bollapragada S, Akbay K, Toledano D, Katlic P, Dulgeroglu O, et al. Automated bed assignments in a complex and dynamic hospital environment. Interfaces 2013 Oct;43(5):435-448. [doi: 10.1287/inte.2013.0701]

20. Guido R, Groccia M, Conforti D. An efficient matheuristic for offline patient-to-bed assignment problems. Eur J Oper Res 2018 Jul;268(2):486-503 [FREE Full text] [doi: 10.1016/j.ejor.2018.02.007]

21. Connelly BL, Certo ST, Ireland RD, Reutzel CR. Signaling theory: a review and assessment. J Manag 2010 Dec 20;37(1):39-67. [doi: 10.1177/0149206310388419]

22. Mavlanova T, Benbunan-Fich R, Koufaris M. Signaling theory and information asymmetry in online commerce. Inf Manage 2012 Jul;49(5):240-247 [FREE Full text] [doi: 10.1016/j.im.2012.05.004]

23. Kazdin AE, Bootzin RR. The token economy: an evaluative review. J Appl Behav Anal 1972;5(3):343-372 [FREE Full text] [doi: 10.1901/jaba.1972.5-343] [Medline: 16795358]

24. Boniecki KA, Moore S. Breaking the silence: using a token economy to reinforce classroom participation. Teach Psychol 2016 Aug 26;30(3):224-227. [doi: 10.1207/s15328023top3003_05]

25. Kim M, Chung J. Sustainable growth and token economy design: the case of steemit. Sustainability 2018 Dec 30;11(1):167. [doi: 10.3390/su11010167]

26. Coates J, Rickwood C, Stacey R. Management Accounting for Strategic and Operational Control, Third Edition. Oxford, United Kingdom: Butterworth-Heinemann; 1996:114-117.

27. Kaplan R, Norton D. The Balanced Scorecard: Translating Strategy into Action. Boston: Harvard Business School Press; 1996.

28. Hayek F. Collectivist Economic Planning. London: Routledge & Kegan Paul Ltd; 1935:211-212.

29.   von Clausewitz CC. On military genius. In: Howard PMP, editor. On War. Princeton, New Jersey: Princeton University Press; 1989:100-112.

30.   Duffield C, Roche M, O'Brien-Pallas L, Catling-Paull C, King M. Staff satisfaction and retention and the role of the nursing unit manager. Collegian 2009;16(1):11-17. [doi: 10.1016/j.colegn.2008.12.004] [Medline: 19388422]

31.   Saari LM, Judge TA. Employee attitudes and job satisfaction. Hum Resour Manage 2004 Nov 17;43(4):395-407. [doi: 10.1002/hrm.20032]

32.   AbuAlRub R, El-Jardali F, Jamal D, Abu Al-Rub N. Exploring the relationship between work environment, job satisfaction, and intent to stay of Jordanian nurses in underserved areas. Appl Nurs Res 2016 Aug;31:19-23. [doi: 10.1016/j.apnr.2015.11.014] [Medline: 27397813]

33.   Spence Laschinger HK, Zhu J, Read E. New nurses' perceptions of professional practice behaviours, quality of care, job satisfaction and career retention. J Nurs Manag 2016 Jul;24(5):656-665. [doi: 10.1111/jonm.12370] [Medline: 26932145]

34.   Sawada Y, Yaginuma S, Tai M, Iyobe S, Mitsuhashi S. Plasmid-mediated penicillin beta-lactamases in pseudomonas aeruginosa. Antimicrob Agents Chemother 1976 Jan;9(1):55-60 [FREE Full text] [doi: 10.1128/AAC.9.1.55] [Medline: 4010]

35.   Panagioti M, Geraghty K, Johnson J, Zhou A, Panagopoulou E, Chew-Graham C, et al. Association between physician burnout and patient safety, professionalism, and patient satisfaction: a systematic review and meta-analysis. JAMA Intern Med 2018 Oct 01;178(10):1317-1330 [FREE Full text] [doi: 10.1001/jamainternmed.2018.3713] [Medline: 30193239]

36.   Gillett G. Autonomy and selfishness. Lancet 2008 Oct 04;372(9645):1214-1215. [doi: 10.1016/s0140-6736(08)61507-x] [Medline: 19105247]

37.   Ivy JW, Meindl JN, Overley E, Robson KM. Token economy: a systematic review of procedural descriptions. Behav Modif 2017 Sep;41(5):708-737. [doi: 10.1177/0145445517699559] [Medline: 28423911]

38.   Daffin LW. Module 9: advanced operant conditioning procedures: consequence focused. In: Principles of Behavior Analysis and Modification. Pullman, WA: Washington State University; 2018:115-141.

39.   Wada Y. Discuss reasons for acute and chronic problems with the crisis of nursing shortage in relation to nursing education and review the strategies and solutions. Nurs Prim Care 2020 Aug 30;4(4):1-4. [doi: 10.33425/2639-9474.1148]

40.   Trendwatch chartbook 2016 the economic contribution of hospitals. American Hospital Asssociation. URL: https://www.aha.org/guidesreports/2016-08-31-chartbook-chart-610-percent-hospital-costs-type-expense-2015-0 [accessed 2019-02-26]

41.   Wunderlich G, Sloan F, Davis C. Staffing and quality of care in hospitals. In: Nursing Staff in Hospitals and Nursing Homes: Is It Adequate?. Washington DC: National Academies Press; 1996.

Original Paper

Public Perception of the Use of Digital Contact-Tracing Tools After the COVID-19 Lockdown: Sentiment Analysis and Opinion Mining

Zhilian Huang[1], PhD; Evonne Tay[1], BSc; Dillon Wee[1]; Huiling Guo[1], MPH; Hannah Yee-Fen Lim[2], PhD; Angela Chow[1,3], PhD

[1]Department of Clinical Epidemiology, Tan Tock Seng Hospital, Singapore, Singapore

[2]Nanyang Business School, Nanyang Technological University, Singapore, Singapore

[3]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

Corresponding Author:
Angela Chow, PhD
Department of Clinical Epidemiology
Tan Tock Seng Hospital
11 Jalan Tan Tock Seng
Singapore, 308433
Singapore
Phone: 65 63577477
Email: Angela_Chow@ttsh.com.sg

Abstract

Background: Singapore's national digital contact-tracing (DCT) tool—TraceTogether—attained an above 70% uptake by December 2020 after a slew of measures. Sentiment analysis can help policymakers to assess public sentiments on the implementation of new policy measures in a short time, but there is a paucity of sentiment analysis studies on the usage of DCT tools.

Objective: We sought to understand the public's knowledge of, concerns with, and sentiments on the use of TraceTogether over time and their preferences for the type of TraceTogether tool.

Methods: We conducted a cross-sectional survey at a large public hospital in Singapore after the COVID-19 lockdown, from July 2020 through February 2021. In total, 4097 respondents aged 21-80 years were sampled proportionately by sex and 4 age groups. The open-ended responses were processed and analyzed using natural language processing tools. We manually corrected the language and logic errors and replaced phrases with words available in the syuzhet sentiment library without altering the original meaning of the phrases. The sentiment scores were computed by summing the scores of all the tokens (phrases split into smaller units) in the phrase. Stopwords (prepositions and connectors) were removed, followed by implementing the bag-of-words model to calculate the bigram and trigram occurrence in the data set. Demographic and time filters were applied to segment the responses.

Results: Respondents' knowledge of and concerns with TraceTogether changed from a focus on contact tracing and Bluetooth activation in July-August 2020 to QR code scanning and location check-ins in January-February 2021. Younger males had the highest TraceTogether uptake (24/40, 60%), while older females had the lowest uptake (8/34, 24%) in the first half of July 2020. This trend was reversed in mid-October after the announcement on mandatory TraceTogether check-ins at public venues. Although their TraceTogether uptake increased over time, older females continued to have lower sentiment scores. The mean sentiment scores were the lowest in January 2021 when the media reported that data collected by TraceTogether were used for criminal investigations. Smartphone apps were initially preferred over tokens, but the preference for the type of TraceTogether tool equalized over time as tokens became accessible to the whole population. The sentiments on token-related comments became more positive as the preference for tokens increased.

Conclusions: The public's knowledge of and concerns with the use of a mandatory DCT tool varied with the national regulations and public communications over time with the evolution of the COVID-19 pandemic. Effective communications tailored to subpopulations and greater transparency in data handling will help allay public concerns with data misuse and improve trust in the authorities. Having alternative forms of the DCT tool can increase the uptake of and positive sentiments on DCT.

XSL•FO
RenderX

Introduction

COVID-19, declared a pandemic by the World Health Organization in March 2020, is highly transmissible, with infections leading to deaths and severe illnesses [1]. Contact tracing has been a critical measure in curbing infectious disease transmissions [2]. However, conventional methods are time-consuming, labor-intensive, subject to recall biases, and unscalable during large-scale outbreaks, such as COVID-19 [3].

Digital contact-tracing (DCT) tools can potentially address the problem of scale during the COVID-19 pandemic by capturing device encounters via Bluetooth-enabled smartphone apps or wearable devices [4,5]. Studies have suggested that DCT tools can help to increase the detection of cases and reduce the time taken for contact tracing by 2.5 times [6-8]. However, a minimum population adoption rate of 60% is required for contact tracing to be effective with DCT tools [9]. At present, Singapore is the only country that has achieved a nationwide DCT adoption rate of more than 70% [10].

Singapore developed a national DCT tool—TraceTogether—in March 2020 and promoted its use after exiting a lockdown in June 2020. Since then, mandatory use of the TraceTogether smartphone app or wearable token for check-ins to enter public venues (such as shopping centers, grocery stores, restaurants, cinemas, schools, and hospitals) has been introduced to increase its adoption [11]. Although the adoption rate of TraceTogether increased from 40% in July to more than 70% in December 2020 [10], the adoption of TraceTogether may have been involuntary under mandatory conditions.

Privacy concerns, lack of trust in the government, and pessimistic views on the effectiveness of DCT tools were barriers against its adoption [12-14]. The plethora of reasons for the hesitancy to adopt DCT tools suggests complex sentiments among users, which may have been deep seated under mandatory conditions. Since the large-scale adoption of DCT tools is unprecedented, understanding the complex sentiments associated with its use would help policymakers to adjust implementation approaches. Traditional qualitative analyses can identify in-depth user perspectives on a topic but are usually confined to smaller samples of text due to their resource-intensive requirements. Natural language processing (NLP) tools are less sensitive in identifying nuances in text data but are able to process a large number of texts in a shorter time frame. Therefore, the use of NLP tools would be most suitable for the analysis of a large number of short texts without in-depth meanings [15].

Opinion mining and sentiment analysis have been performed widely to understand the public's reaction and challenges faced during the COVID-19 pandemic [16,17]. These methods are useful for consolidating a large amount of information in a short period. For example, policymakers can study public sentiments on new events or COVID-19 measures and tailor public health communications to mitigate negative emotions arising from the event or measure [17]. An increasing number of studies have utilized short social media texts, such as Tweeter posts, to understand public sentiments on the COVID-19 pandemic [18,19]. Despite the benefits of opinion mining and sentiment analysis in rapidly garnering the public's opinion, the data collected from online platforms, such as microblogs, social media, app store reviews, and online surveys, were biased toward the more privileged and technologically savvy individuals [20,21]. Overrelying on data collected from social media sites would omit the views of social groups that do not use these sites [22,23].

There is also a paucity of sentiment analysis studies on the usage of DCT tools. An Irish study analyzed the sentiments of DCT app reviews and found predominantly positive sentiments; however, the review focused on voluntary app users who may be more accepting of DCT tools [20]. Given the lack of studies on user perspectives on DCT tools and biases of the data sources used for opinion mining, we sought to understand the public's knowledge of, concerns with, and sentiments on the use of DCT tools over time, as mandated by the authorities for pandemic control, across an extensive demographic and age distribution.

Methods

Study Design

We conducted a cross-sectional survey in the 2 busiest ambulatory clinics at the second-largest public hospital in Singapore, starting from 1 month after the nationwide COVID-19 lockdown in 2020. Data collection occurred over 8 months from July 2020 through February 2021 during patients' or their caregivers' visit to the clinic. Respondents from ages 21 to 80 years were sampled proportionally by sex and four 15-year age groups to cover the perspectives of digital natives and digital immigrants. We included only citizens and permanent residents of Singapore as this population would best fit the context of our study.

Timeline of TraceTogether Events in Singapore

The use of TraceTogether was widely promoted after the COVID-19 lockdown in Singapore. The smartphone app was initially promoted to trace encounters with users in close proximity but was updated in June 2020 to collect personal identifiers for more effective contact tracing. The app is available in multiple languages, including Bengali, Burmese, Chinese, English, Hindi, Melayu, Tamil, and Thai. In July 2020, the token form of TraceTogether was made available to seniors who do not own smartphones. Each token weighs 15 g, with dimensions that are 62 mm long, 15 mm thick, and 45mm wide. From September to November 2020, the government made a series of announcements to promote the uptake of TraceTogether. All Singapore citizens were eligible to collect a free TraceTogether token to facilitate mandatory safe entry check-ins at all public venues [11], and social restrictions would

be eased further if at least 70% of the population adopted TraceTogether [24]. In early January 2021, the Singapore police force used the TraceTogether data for criminal investigations under the Criminal Procedure Code (CPC) [25]. Clarifications were made when the act evoked a public outcry on personal data protection.

Survey Instrument

We designed a 14-item survey questionnaire based on literature review and included questions on the status of digital device usage and willingness to use TraceTogether (pre- and postsharing on DCT tools). Three open-ended questions were used to determine respondents' knowledge of TraceTogether, top three concerns with any DCT technology, and the reasons for their preference for the form factor of TraceTogether (refer to Multimedia Appendix 1 for the questionnaire).

Data Collection

We trained all data collectors to ensure that the questionnaire was appropriately administered by the interviewer. Information on respondents' perceptions of a DCT tool was collected using TraceTogether as an example. We then provided a 2-minute explanation on the purpose of DCT tools at the end of the survey and asked respondents again whether they would be willing to use a DCT tool and, if so, whether they preferred an app or a token and the reason for their choice. Demographic information was collected to perform segmented analyses.

This study focused on the open-ended responses from respondents, which included asking the respondents their thoughts on the purpose, data security, and usage of TraceTogether and their concerns with TraceTogether. The reasons for the choice of a smartphone app or token were also analyzed.

Descriptive Statistics

We classified respondents into 4 age and sex categories (younger females, older females, younger males, and older males) and classified those who were above the age of 50 years as older adults and those 50 or under 50 years as younger adults. Mean and SDs were computed for age, while proportions were computed for other categorical variables, such as demographics, smartphone ownership, awareness of TraceTogether, willingness to use TraceTogether, and preference for its form factor. We also presented the bimonthly uptake rate of TraceTogether by age and sex to show the impact of the policy measures to boost uptake rates.

Data Processing

The open-ended responses were manually processed to correct language and spelling errors. Abbreviations were written in full, and the informal and colloquial form of the English language was rephrased to the formal form. For example, "don't like" was rephrased as "dislike" as "don't" would likely be removed as a stop word, while "like" would be detected as a positive sentiment, although the phrase implies a negative sentiment. Important phrases on "knowledge of TraceTogether" were standardized 3-word phrases and analyzed as trigrams (refer to Table S1 in Multimedia Appendix 1), while other sections were analyzed as bigrams. The processed trigrams were subsequently replaced with a phrase that was closer to their original meaning before it was presented graphically (eg, the pre-processed phrase "Location unknown uncollected" was replaced with "Location data NOT collected" when presented graphically).

The preprocessed responses were then processed with NLP tools. All phrases were tokenized (split into smaller units), and the sentiment score of each phrase was computed by summing the sentiment scores of all the tokens in the phrase. Stopwords, such as prepositions and connectors, were removed according to the *stopwords* package in R, followed by implementation of the bag-of-words model (simplifying the representation of words) to calculate the occurrence of bigrams and trigrams in the data set. Demographic and time filters were applied to segment the responses, as required (Figure 1).

Figure 1. Process of data processing and analysis. *Refer to Table S1 and Figure S1 in Multimedia Appendix 1.

```
┌─────────────────────────────────┐
│   Text collected from 4097 patient │
│            interviews              │
└─────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────┐
│  Text manually cleaned to remove  │
│  language errors and rephrased for │
│            consistency             │
│                                    │
│   (eg, not convenient=inconvenient)* │
└─────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────┐
│     Text document tokenization    │
└─────────────────────────────────┘


┌─────────────────────────────────┐
│   Sentiment score calculated for each │
│             document               │
└─────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────┐
│        Stopwords removed          │
└─────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────┐
│ Demographic filters applied to segment │
│              groups                │
└─────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────┐
│   Bag-of-words model implemented to │
│  calculate the occurrence of bigrams │
└─────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────┐
│  Original data reviewed based on top few │
│       bigrams of each category     │
└─────────────────────────────────┘
```

Sentiment Analysis

We used the *syuzhet* package for sentiment analysis as it incorporates 3 other lexicons developed by other groups [26]. The *bing* lexicon contains a list of words classified into positive and negative sentiments, while the *nrc* lexicon classifies words into 8 other emotions on top of the positive and negative sentiments. The *afinn* and *syuzhet* lexicons contain a database of words with a sentiment score. We compared the sentiment scores derived from the *syuzhet* and *afinn* lexicon and did not find substantial differences in sentiment patterns [26,27]. Hence, we used the *syuzhet* lexicon for our analyses as the database contains a wider range of vocabulary compared to the *afinn* database. Each word in the *syuzhet* library has a value of between –1 and 1. Words with positive connotations are scored positively, while those with negative connotations are scored negatively. The sentiment score of a response statement was computed by summing the values of all the words in the

statement that could be found in the *syuzhet* library. All analyses were performed using RStudio version 1.2.5033.

Some of the respondents would comment on the disadvantages of the TraceTogether token when asked why they preferred the smartphone app or vice versa. Hence, the reasons for respondents' preference for either the TraceTogether tool (smartphone app or portable token) were split into token- or smartphone-related comments before sentiment analysis.

Ethical Considerations

This study was approved by the National Healthcare Group (NHG) Domain Specific Review Board (DSRB) in Singapore (NHG DSRB ref. 2020/00775). A waiver of written informed consent was granted, and implied consent was assumed if the individual agreed to respond to the survey.

Results

Recruitment Rate and Demographics of Respondents

We approached 6260 potential respondents and excluded 744 (11.88%) who did not meet the inclusion criteria. Of 5229 eligible participants, we interviewed 4097 (78.35%) respondents in total. Approximately a quarter of the respondents who were interviewed declined to respond to the open-ended questions.

Table 1 shows the demographics of respondents. Age and sex were proportionately sampled during data collection. Hence, respondents were divided into 4 age and sex categories. The Chinese race was slightly oversampled in older adults as the Chinese race constitutes about 76% of the Singapore population. A higher proportion of younger adults were tertiary educated compared with older adults, but the overall education level of respondents was representative of the general population. A smaller proportion of older adults were employed, as 796 (38.81%) of 2051 older adult respondents were retirees.

Table 1. Baseline characteristics of respondents (N=4097).

Characteristics	Total respondents	Younger males	Younger females	Older males	Older females
Age (years), mean (SD)	50.2 (16.8)	35.4 (8.7)	35.7 (8.5)	64.8 (7.9)	64.7 (7.9)
Ethnicity, n/N (%)					
Chinese	3330/4097 (81.27)	802/1024 (78.32)	756/1022 (73.97)	867/1027 (84.42)	905/1024 (88.38)
Malay	315/4097 (7.69)	107/1024 (10.45)	148/1022 (14.48)	49/1027 (4.77)	50/1024 (4.88)
Indian	354/4097 (8.64)	84/1024 (8.20)	84/1022 (8.22)	91/1027 (8.86)	56/1024 (5.47)
Others	98/4097 (2.39)	31/1024 (3.03)	34/1022 (3.33)	20/1027 (1.95)	13/1024 (1.27)
Education level, n/N (%)					
Tertiary	1296/4097 (31.63)	488/1024 (47.66)	489/1022 (47.85)	214/1027 (20.84)	105/1024 (10.25)
Employment status[a], n/N (%)					
Employed	2728/4097 (66.59)	870/1024 (84.96)	872/1022 (85.32)	548/1027 (53.36)	438/1024 (42.77)
Digital device usage[b], n/N (%)					
Owns a smartphone	3712/4097 (90.60)	1022/1024 (99.80)	1019/1022 (99.71)	871/1027 (84.81)	800/1024 (78.13)
Heard of TraceTogether	3818/4097 (93.19)	987/1024 (96.39)	972/1022 (95.11)	934/1027 (90.94)	925/1024 (90.33)
Willing to use TraceTogether (presharing)	3143/4097 (76.71)	739/1024 (72.17)	767/1022 (75.05)	806/1027 (78.48)	831/1024 (81.15)
Willing to use TraceTogether (postsharing; 36 [0.9%] missing)	3674/4061 (90.47)	908/1021 (88.93)	924/1018 (90.77)	903/1010 (89.41)	939/1012 (92.79)
Form factor preference (willing to use TraceTogether postsharing; 127 [3.5%] missing), n/N (%)					
Smartphone app	1900/3547 (53.56)	604/872 (69.27)	602/887 (67.87)	399/872 (45.76)	295/916 (32.21)
Token	1647/3547 (46.43)	268/872 (30.73)	285/887 (32.13)	473/872 (54.24)	621/916 (67.79)

[a]Full-time, part-time, self-, or temporary employment.

[b]Uptake of TraceTogether was not presented in this table, as uptake rates changed over time with the policy measures. Refer to Figure 2 for bimonthly uptake rates.

Figure 2. The proportion of TraceTogether uptake and mean (syuzhet) sentiment score of concerns with TraceTogether over time.
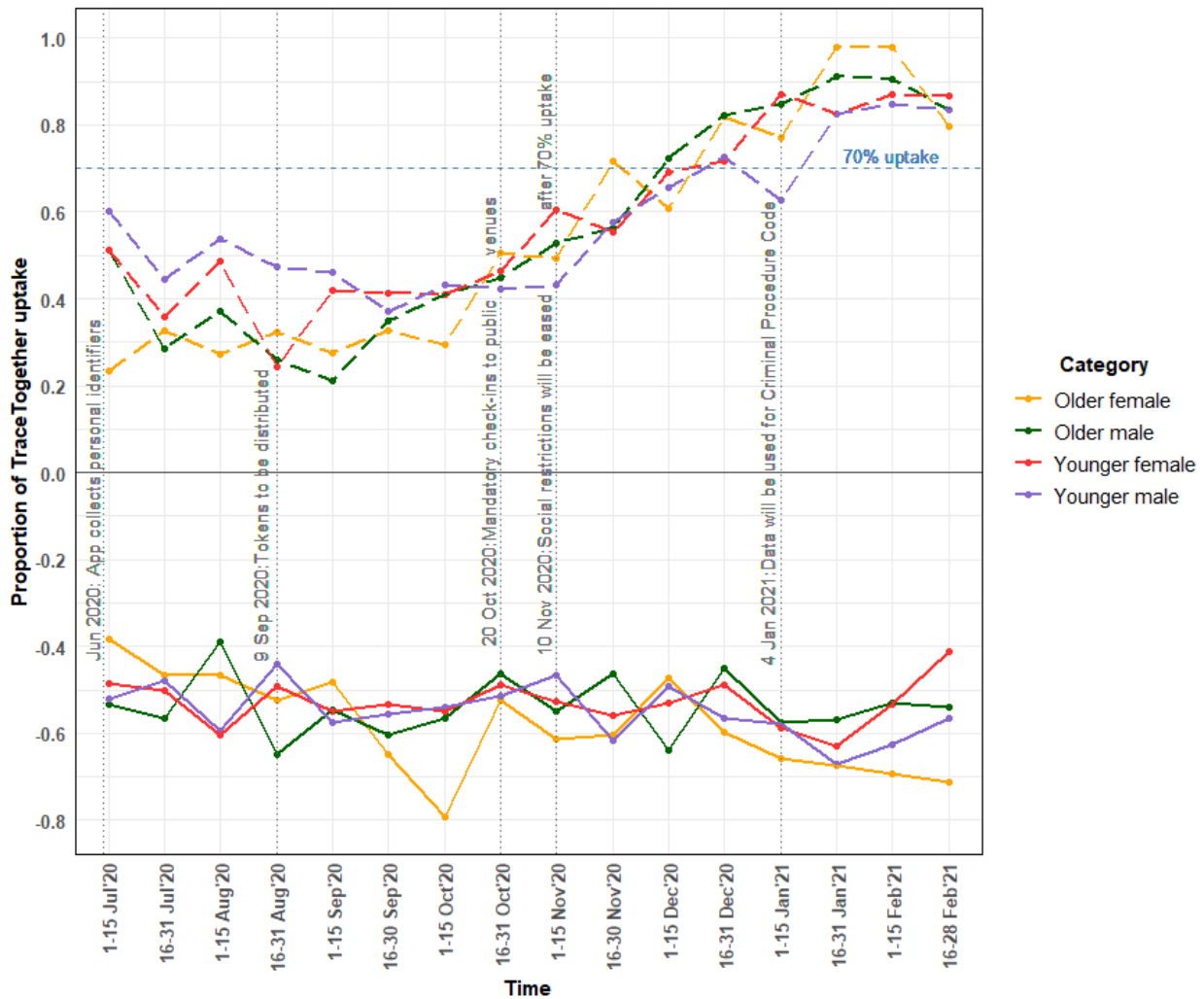
The overall smartphone ownership was 3712 (90.60%) of 4097 people. However, the proportion of older adults who owned a smartphone was lower than that of younger adults. Older females had the lowest smartphone ownership compared with other groups. The majority of respondents (3818/4097, 93.19%) had heard of TraceTogether at the time of the survey. The willingness to use TraceTogether increased across all age and sex categories after the study team explained the rationale and benefits of TraceTogether to the respondent.

Knowledge of TraceTogether

We counted the occurrence of unique trigrams and further collapsed trigrams with similar meanings to reduce the number of statements. Figure 3 shows the proportion of top trigrams classified into 2-month periods. All but 1 period had trigrams covering at least 70% of responses. Overall, the proportion of trigrams, representing respondents' knowledge and perceptions of TraceTogether, changed over time.

The top 6 trigrams from respondents' opinion on the purpose of TraceTogether ("What do you think TraceTogether is for?") were "Contact-tracing purpose/trace close-proximity contacts," "Location-tracing purpose," "Location-tracking purpose,"

"COVID-19-positive patient," "Receive alert notification," and "Scan QR code/location check-ins." The proportion of mentions on contact tracing decreased (from 414/1806 [22.92%] to 365/2755 [13.25%]), while mentions on QR code scanning and location check-ins increased (from 19/588 [3.23%] to 133/1362 [9.77%]) over time.

The top 6 trigrams from respondents' opinion on the usage of TraceTogether ("What do you think users need to do?") were "Activate Bluetooth setting," "Activate GPS tracker," "Activate mobile data," "Activate/download phone app," "Carry token alongside," and "Scan QR code/scan NRIC/location check-ins" (where NRIC stands for National Registration Identity Card). Initially, respondents thought they had to download the app (41/292, 14.04%) and activate Bluetooth on their smartphones (174/292, 59.59%) to use TraceTogether. Over time, the proportion of mentions on Bluetooth activation decreased from 174 (59.59%) of 292 to 284 (29.71%) of 956, while the proportion of mentions on scanning QR codes/location check-ins increased from 9 (3.08%) of 292 to 301 (31.49%) of 956. The proportion of mentions on the need to carry the TraceTogether token also increased from 0 in August 2020 to 81 (8.47%) of 956 in December 2020.

Figure 3. Participants' responses on their knowledge of TraceTogether changed over time. The occurrence of trigrams and their proportions relative to all trigrams in a time category were computed. GPS: Global Positioning System; NRIC: National Registration Identity Card. *Trigrams were rephrased for clarity.
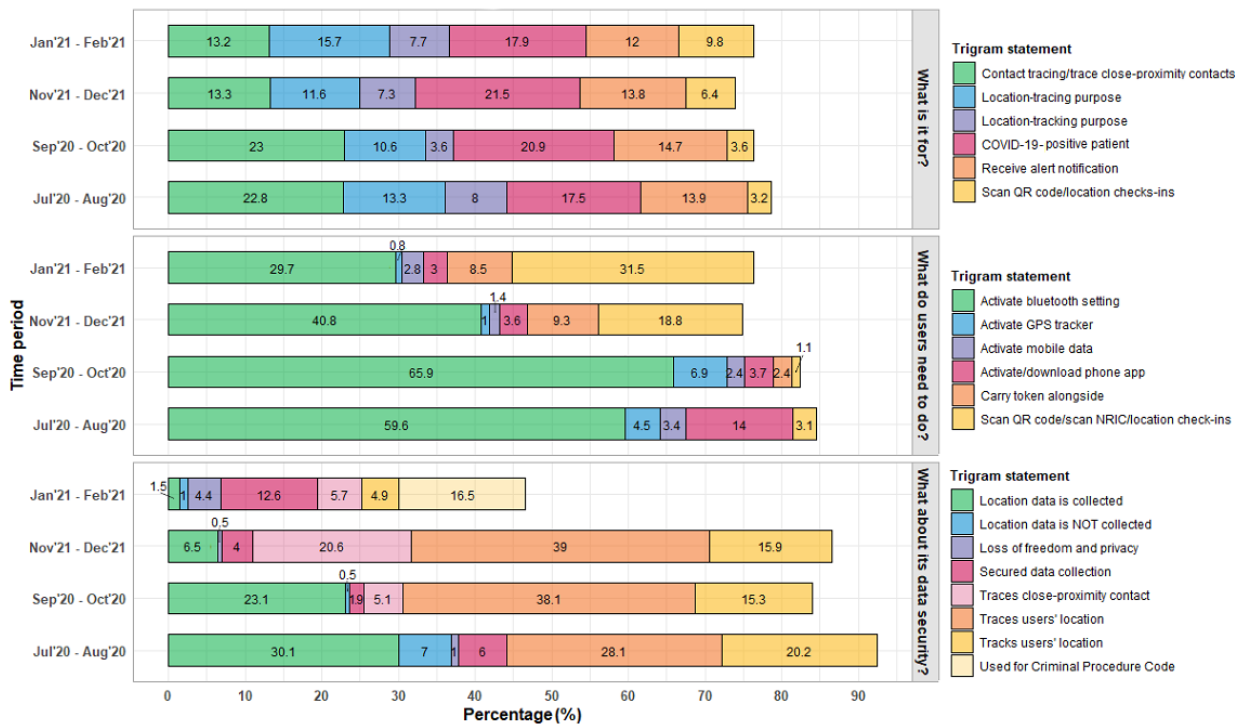
When respondents were asked their opinion on the data security of TraceTogether, three-quarters of the responses from July to October 2020 were mentions of "location data collected" or "users' location traced/tracked." By December 2020, the proportion of mentions related to location tracking/tracing/data collection decreased to 263 (61%) of 428 and subsequently to 25 (6.44%) of 388 by February 2021. The proportion of mentions on "secured data collection" doubled from 18 (5.96%) of 302 in July-August 2020 to 49 (12.63%) of 388 in November-December 2020. However, 64 (16.49%) of the 388 trigrams in January-February 2021 were mentions of the use of data collected by TraceTogether for the "Criminal Procedure Code."

Concerns With the Use of TraceTogether

We present the proportion of TraceTogether uptake (question item on "Are you currently using the TraceTogether app or Token?") and the mean *syuzhet* sentiment score of concerns with TraceTogether (open-ended question on "Please list your top three [3] main concerns with any DCT technology [not limited to TraceTogether]") at 2-week intervals in Figure 2. The plots were segmented into 4 categories: older males, older females, younger males, and younger females. Older adults were aged between 51 years and 80 years, while younger adults were between 21 years and 50 years of age. Uptake rates increased rapidly after the announcement of mandatory TraceTogether check-ins at public venues in mid-October 2020 and reached 70% by December 2020. The magnitude of the sentiment scores was used to compare sentiment changes over

time, and the concerns with TraceTogether were negative overall.

Younger males had the highest TraceTogether uptake (24/40, 60%), while older females had the lowest uptake (8/34, 24%) in the first half of July 2020. This trend was reversed in mid-October after the announcement on mandatory TraceTogether check-ins at public venues. In mid-February 2021, the TraceTogether uptake of older adults surpassed 90% (93/99), while that of younger adults was 80%-90% (100/116).

The mean sentiment scores were the lowest in January 2021 when the media reported that the data collected by TraceTogether were used for criminal investigations. Older females also had decreased sentiment scores as their TraceTogether uptake increased over time. This group of respondents were mainly concerned about data breaches, privacy violation, and the pressure to adopt a new technology that they were unfamiliar with. Four respondents had misconceptions about the Bluetooth technology and cited health concerns about possible radiation emitted by TraceTogether. Younger adults had similar concerns with privacy violation but were also concerned about Bluetooth battery consumption on their smartphones.

We present the occurrence of bigrams on respondents' concerns with the usage of TraceTogether in Table 2. The bigrams were classified into 5 categories. Each bigram was presented with a corresponding example of a response statement and the sentiment score of the statement.

Table 2. Top bigrams (N=3995) of respondents' concerns with TraceTogether.

Bigram	Occurrence, n (%)	Example of response statement	Sentiment score of example statement
TraceTogether app inconvenience			
Battery drainage	489 (12.24)	*Bluetooth needs to be activated, which causes battery drainage and slows down the app.* [Male, 57 years, September 2020]	–0.25
Technical glitch	83 (2.08)	*Technical glitch. It malfunctions on my phone all the time and does not seem to capture any interactions that I made with the people around me. That is why I stopped utilizing it.* [Female, 41 years, August 2020]	–1.5
Phone battery	77 (1.93)	*Inconvenient. If [my] phone battery [goes] flat, it [TraceTogether] would not work anymore.* [Female, 57 years, November 2020]	–0.25
User unfriendly	22 (0.55)	*App is user unfriendly.* [Female, 49 years, October 2020]	–0.5
Bluetooth battery	22 (0.55)	*Bluetooth battery drainage.* [Respondents aged 21-71 years, both sexes, all months]	–0.25
Memory space	10 (0.25)	*Lack of phone memory storage space.* [Female, 27 years, July 2020]	–0.75
Phone memory	9 (0.23)	*Lack of phone memory storage space.* [Female, 27 years, July 2020]	–0.75
Language barrier	5 (0.13)	*Language barrier. [The] Interface is in English. The first page should especially be in mandarin as 70% of the population is Chinese and some may not understand [English].* [Male, 68 years, July 2020]	–0.5
TraceTogether app and token data security			
Privacy violation	386 (9.66)	*Data privacy violation. Even though they [the authorities] say they collect minimal data, you would not know if they changed it down the road since nobody reads the terms and conditions.* [Male, 46 years, November 2020]	–0.5
Data privacy	238 (5.96)	*Location data privacy violation. You do not want others to know some places you go to.* [Female, 68 years, January 2021]	–0.5
Dislike location	139 (3.48)	*Linked with location tracking and location tracing*	__a
Location tracking	96 (2.40)	*Dislike location tracking. The app will track my location like an ankle collar on criminals.* [Male, 54 years, December 2020]	–1.5
Location tracing	53 (1.33)	*Dislike location tracing. Mistrust the government on data safety.* [Female, 57 years, November 2020]	–1.2
Data insecurity	81 (2.03)	*Lack of transparency on data usage, concerned about data insecurity.* [Male, 29 years, July 2020]	–2.1
Data leak	66 (1.65)	*Afraid of [data] being hacked and having data leak.* [Male, 27 years, September 2020]	–2.1
Data unprotected	9 (0.23)	*Data [are] unprotected. Dislike having to input IC [identification] number. Phone number should suffice.* [Male, 68 years, September 2020]	–1.8
TraceTogether app and token data misuse			
PDPA[b] violation	152 (3.80)	*PDPA violation. Dislike that personal details have to be entered before you can utilize it.* [Female, 42 years, December 2020]	–1.5
Data breach	148 (3.70)	*Who is responsible if there is a data breach and how is the breach being handled?* [Female, 50 years, July 2020]	–0.5
Privacy invasion	41 (1.03)	*Privacy invasion. A sense that you are being stalked.* [Male, 67 years, July 2020]	–1.4
Jeopardize bank details	17 (0.43)	*Jeopardize bank and credit card details.* [Respondents aged 26-67 years, both sexes, September 2020-February 2021]	–0.5
Credit card details	16 (0.40)	*Jeopardize bank and credit card details.* [Respondents aged 26-67 years, both sexes, September 2020-February 2021]	–0.5
Personal information	18 (0.45)	*Privacy violation as there may be a personal information leak.* [Female, 47 years, January 2021]	–0.85
Information leak	12 (0.30)	*Privacy violation as there may be a personal information leak.* [Female, 47 years, January 2021]	–0.85

XSL•FO
RenderX

Bigram	Occurrence, n (%)	Example of response statement	Sentiment score of example statement
TraceTogether app and token efficiency and efficacy of contact tracing			
Contact tracing	58 (1.45)	*Even with contact tracing we are unable to immediately know who is infected or if I met them, so it is inefficient.* [Female, 40 years, February 2021]	–1.35
TraceTogether inaccurate	6 (0.15)	*TraceTogether [is] inaccurate. It is pointless if I utilize and others do not.* [Male, 50 years, October 2020]	–1.5
Delayed notification	6 (0.15)	*The app notified me of possible exposure on the 14th, but I received delayed notification only on the 30th.* [Male, 30 years, December 2020]	–0.15
TraceTogether token design			
Size cumbersome	35 (0.88)	*The token is badly designed, has a limited lifespan and its large size [is] cumbersome.* [Male, 74 years, December 2020]	–2
Token size	12 (0.30)	*Cumbersome token size and it is unaccepted at some stores.* [Female, 51 years, December 2020]	–0.5

[a]Not applicable.

[b]PDPA: Personal Data Protection Act.

Inconvenience Created by the TraceTogether App

In this study, 717 (17.95%) of 3995 bigrams were concerns with the inconvenience created by the TraceTogether app. The concerns included smartphone battery drainage due to Bluetooth activation, frustration with technical glitches, and the app taking up phone memory space. Older individuals may have language barriers with using the app, and individuals with dependents may find the process of checking into locations troublesome.

> *Bluetooth needs to be activated, which causes battery drainage and slows down the app.*

> *Inconvenient. If [my] phone battery [goes] flat, it [TraceTogether] would not work anymore.*

> *User unfriendly. Will appreciate [it] if you can include your child in the parents' app as well so can save time on scanning as the child's details are inside the parents' app.*

> *Language barrier. [The] interface is in English. The first page should especially be in [M]andarin as 70% of the population is Chinese and some may not understand [English].*

Concerns With Data Security in the TraceTogether App and Token

In this study, 1068 (26.73%) of 3995 bigrams were concerns with the data security of TraceTogether. Respondents disliked location tracking or tracing and felt that their (data) privacy was/will be violated with the use of TraceTogether. Respondents also felt that data transparency was insufficient, and they were insecure about data leaks should they lose their token.

> *Data privacy violation. Even though they [the authorities] say they collect minimal data, you would not know if they changed it down the road since nobody reads the terms and conditions.*

> *Location data privacy violation. You do not want others to know some places you go to.*

> *Dislike location tracking. The app will track my location like an ankle collar on criminals.*

> *Dislike location tracing. Mistrust the government on data safety.*

> *Lack of transparency on data usage, concerned about data insecurity.*

Concerns With Data Misuse by the TraceTogether App and Token

In this study, 404 (10.11%) of 3995 bigrams were concerns with data misuse, such as violation of the Personal Data Protection Act (PDPA), data breaches, and leakage of personal information and credit card details. Respondents also felt that tagging the TraceTogether device provided a sense of privacy invasion and insecurity about possible data breaches.

> *PDPA violation. Dislike that personal details have to be entered before you can utilize it.*

> *Who is responsible if there is a data breach and how is the breach being handled.*

> *Privacy invasion. A sense that you are being stalked.*

Concerns With the Efficiency and Efficacy of Contact Tracing

In this study, 70 (1.75%) of 3995 bigrams were mentions of the accuracy of TraceTogether and delayed notifications. Respondents were concerned that TraceTogether would be inaccurate if most of the population does not utilize TraceTogether appropriately. Some respondents mentioned that they did not receive timely app notifications of possible exposures.

> *Inefficiency of the tool. The success of contact tracing depends on the cooperation of citizens.*

> *The app notified me of possible exposure on the 14th (of the month), but I received delayed notification only on the 30th (of the month).*

Dissatisfaction With the TraceTogether Token Design

In this study, 47 (1.18%) of 3995 bigrams were mentions of dissatisfactions with the TraceTogether token form factor. Respondents felt that the large token size makes it cumbersome

to carry around. Other issues related to the token include its limited battery span, unsightly aesthetics, and inability to check in at smaller stores without token scanners.

> *The token is badly designed, has a limited lifespan, and its large size [is] cumbersome.*
>
> *Cumbersome token size and it is unaccepted at some stores.*

Preference for the TraceTogether Tool

Respondents' preference for the type of the TraceTogether tool (smartphone app or portable token) and sentiment scores of the reason for their preferred type over time are shown in Figure 4. In the first 2 months of data collection, in July and August 2020, two-thirds (96/150, 64%) and three-quarters (161/217, 74.2%) of respondents preferred the smartphone app over the token. Over time, the preference for the type of TraceTogether tool equalized among respondents as tokens became accessible to the whole population.

Figure 4. The proportion of respondents' preference for the TraceTogether tool (smartphone app or token) and the sentiment scores of the reason for their choice over time. Note: The proportion of TraceTogether preferences are based on cross-sectional time series data.
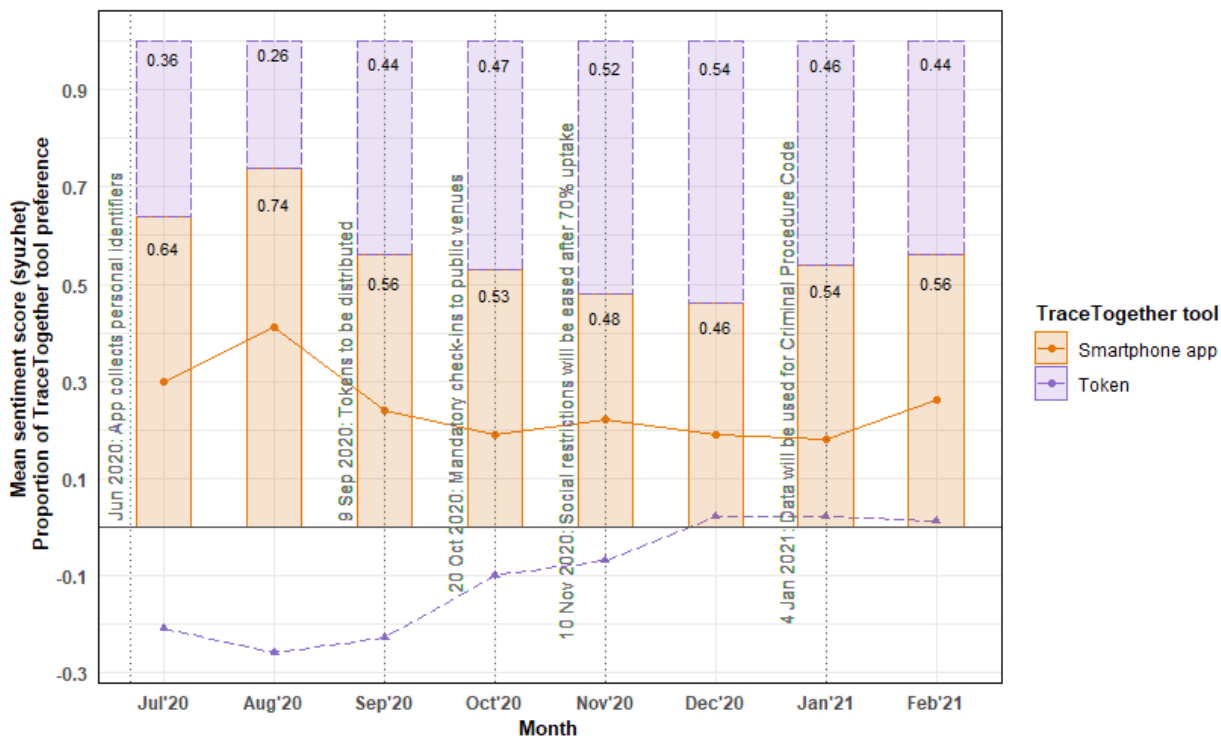
The sentiment scores of the reason for the preferred type of TraceTogether tool moved in tandem with the proportion of respondents' indicating preference for that particular type. The sentiments on token-related comments became more positive as the preference for tokens increased. Similarly, the sentiments on smartphone app–related comments became less positive as the preference for the smartphone app decreased. Overall, respondents had more positive sentiments on the use of the TraceTogether app compared with the token.

Respondents preferring the smartphone app felt that the app was a convenient option since they would always have their phones with them. Respondents preferring the app also felt burdensome to have to remember to bring the token when going out and were worried that the token, if misplaced, would be misused by the finder. In addition, some respondents commented that the plastic used to manufacture the tokens was environmentally unfriendly and that the tokens were unsightly and bulky to carry. There were also smaller shops that did not allow the checking-in of tokens.

Respondents preferring the token felt that the tokens were suitable for the elderly with difficulty using smartphone apps.

They liked that the token does not consume the smartphone's battery and mobile data and that there is no need to charge the token or worry about their smartphone battery going flat.

Discussion

Principal Findings

We found differences in the sentiments of respondents across age, sex, and time under voluntary and mandatory use of TraceTogether. The application of NLP techniques on unstructured free-text responses to open-ended questions in an interviewer-administered questionnaire implemented consistently over 8 months enabled us to quantify and examine the changes in public sentiments as the COVID-19 pandemic evolved. Such analyses will provide useful and timely feedback to policymakers on the impact of public health policies and measures imposed on the population and enable them to fine-tune them for greater public acceptance and compliance.

The knowledge of TraceTogether did not differ across age groups. Respondents' knowledge of the purpose of TraceTogether did not change substantially over time, except for a slight shift from its use for contact tracing to scanning of

QR codes for location check-ins. A small proportion of respondents (450/5206, 8.64%) had the misconception that TraceTogether tracks their location. Polls have shown that 1 reason for the hesitancy in TraceTogether uptake is the misconception that TraceTogether tracks the user's location [28]. More could be done by the authorities to address such misconceptions in public education.

Respondents' perception of the security of the data collected by TraceTogether shifted from being location focused from July to December 2020 to a focus on the Criminal Procedure Code and loss of freedom and privacy in January-February 2021. The overall sentiments on the concerns with TraceTogether were most negative in January 2021. Despite the negative sentiments on the usage of TraceTogether data for the CPC, there was also a higher proportion of mentions of secured data collection in January-February 2021, implying improved knowledge of and confidence in TraceTogether's data handling among some respondents. The media is powerful in eliciting knee-jerk reactions among the public. Although such reactions may be short lived, it is imperative to promptly address any public concern to prevent long-term repercussions [29].

The sentiments on concerns with TraceTogether became more negative after the government's announcement on the mandatory use of TraceTogether for check-ins to enter public venues, although the reported uptake rate of TraceTogether increased. This observation suggests involuntary uptake of TraceTogether required for social interactions to resume. In another study, we assessed the trade-offs of social interactions and incentives on the use of DCT tools and found that most people would prefer to use a DCT tool in exchange for more social interactions under conditions of social restrictions during the COVID-19 pandemic.

The involuntary uptake of TraceTogether may lead to negative sentiments due to a lack of understanding of the data transparency and the need for users to tag their identifiers with the tool. Although the negative sentiments could be transient and existent only during the COVID-19 pandemic, prolonged negative population sentiments may lead to future political repercussions if the benefits of the mandatory measures are unappreciated [30].

The sentiments and preference for TraceTogether tokens improved after mass token distributions to the public. Although smartphone apps were preferred, having an alternative type of TraceTogether tool could have improved the overall uptake of TraceTogether. Reducing barriers to accessibility may have helped to increase the TraceTogether uptake rate as no other country has nationally distributed alternative forms of the DCT tool or achieved more than 70% uptake [31].

Limitations

There are various limitations to this study. First, the observations were cross-sectional, as we did not assess opinion changes of the same respondents over time. Nonetheless, the serial cross-sectional surveys on proportionately sampled respondents with a good representation of age and sex at every period provided invaluable insights into the changes in the population's opinions over time. Second, the *stopwords* package removed too many words, while the existing sentiment libraries did not have sentiment scorings for colloquial phrases such as "don't like," "don't want," and "not friendly." We had to manually replace these words with words found in the sentiment library to apply a sentiment score to the phrase. Regardless, the meaning of the words was retained. Third, manual data cleaning is time-consuming and may not be feasible for analyzing a large data set in a short amount of time. Lastly, respondents were patients and visitors of a public hospital and may not be representative of the Singapore population. However, a good representation of sex and age groups were sampled.

Future studies could explore crowdsourcing to develop a sentiment library that better suits the local context to reduce the time spent on data preprocessing. A stop word library that excludes (does not remove) words used in colloquial phrases will also reduce efforts on data preprocessing. Timely awareness of the public's sentiments on a new policy will allow policymakers to adjust their approach to public communication [32]. The insights gained from subpopulations, such as the elderly and adults who are not technologically astute, provide opportunities to tailor interventions that can help them to better adapt to the new technology.

Conclusion

In conclusion, the public's knowledge of and concerns with using a mandatory DCT tool varied with the national regulations and public communications over time with the evolution of the COVID-19 pandemic. Effective communications tailored to subpopulations and greater transparency in data handling will help allay public concerns with data misuse and improve trust in the authorities. Having alternative forms of the DCT tool can increase the uptake of and positive sentiments on DCT tools.

Authors' Contributions

ZH conceived the manuscript, analyzed the data, and drafted the manuscript. ET and DW assisted in analyzing the data. HYFL and AC provided support for the study and study planning. All authors critically reviewed the manuscript and approved the final version of the manuscript prior to submission.

XSL•FO

RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional information on data processing and questionnaire.
[DOCX File , 63 KB - formative_v6i3e33314_app1.docx]

References

1. Jin Y, Yang H, Ji W, Wu W, Chen S, Zhang W, et al. Virology, epidemiology, pathogenesis, and control of COVID-19. Viruses 2020 Mar 27;12(4):372 [FREE Full text] [doi: 10.3390/v12040372] [Medline: 32230900]
2. Contact Tracing in the Context of COVID-19. COVID-19: Surveilllance, Case Investigation and Epidemiological Protocols. URL: https://www.who.int/publications/i/item/contact-tracing-in-the-context-of-covid-19 [accessed 2022-02-24]
3. Kretzschmar M, Rozhnova G, Bootsma M, van Boven M, van de Wijgert JHHM, Bonten M. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. Lancet Public Health 2020 Aug;5(8):e452-e459 [FREE Full text] [doi: 10.1016/s2468-2667(20)30157-2]
4. Huang Z, Guo H, Lee Y, Ho EC, Ang H, Chow A. Performance of digital contact tracing tools for COVID-19 response in Singapore: cross-sectional study. JMIR Mhealth Uhealth 2020 Oct 29;8(10):e23148 [FREE Full text] [doi: 10.2196/23148] [Medline: 33006944]
5. Anglemyer A, Moore T, Parker L, Chambers T, Grady A, Chiu K. Digital contact tracing technologies in epidemics: a rapid review. Cochrane Database Syst Rev 2020;18(8):CD013699. [doi: 10.1002/14651858.cd013699]
6. Barrat A, Cattuto C, Kivelä M, Lehmann S, Saramäki J. Effect of manual and digital contact tracing on COVID-19 outbreaks: a study on empirical contact data. J R Soc Interface 2021 May;18(178):20201000 [FREE Full text] [doi: 10.1098/rsif.2020.1000] [Medline: 33947224]
7. Rodríguez P, Graña S, Alvarez-León EE, Battaglini M, Darias FJ, Hernán MA, RadarCovidPilot Group, et al. A population-based controlled experiment assessing the epidemiological impact of digital contact tracing. Nat Commun 2021 Jan 26;12(1):587 [FREE Full text] [doi: 10.1038/s41467-020-20817-6] [Medline: 33500407]
8. Ministry of Health. TraceTogether and SafeEntry to Be Enhanced in Preparation for Further Opening of the Economy. URL: https://www.moh.gov.sg/news-highlights/details/tracetogether-and-safeentry-to-be-enhanced-in-preparation-for-further-opening-of-the-economy [accessed 2022-02-24]
9. Baraniuk C. Covid-19 contact tracing: a briefing. BMJ 2020 May 13;369:m1859. [doi: 10.1136/bmj.m1859] [Medline: 32404304]
10. Huang Z, Guo H, Lim HY, Chow A. Awareness, acceptance, and adoption of the national digital contact tracing tool post COVID-19 lockdown among visitors to a public hospital in Singapore. Clin Microbiol Infect 2021 Jul;27(7):1046-1048 [FREE Full text] [doi: 10.1016/j.cmi.2021.01.007] [Medline: 33482353]
11. Wong L. Use of TraceTogether App or Token Mandatory by End Dec. URL: https://www.straitstimes.com/singapore/use-of-tracetogether-app-or-token-mandatory-by-end-dec [accessed 2020-11-16]
12. O'Callaghan ME, Buckley J, Fitzgerald B, Johnson K, Laffey J, McNicholas B, et al. A national survey of attitudes to COVID-19 digital contact tracing in the Republic of Ireland. Ir J Med Sci 2021 Aug 16;190(3):863-887 [FREE Full text] [doi: 10.1007/s11845-020-02389-y] [Medline: 33063226]
13. Megnin-Viggars O, Carter P, Melendez-Torres GJ, Weston D, Rubin GJ. Facilitators and barriers to engagement with contact tracing during infectious disease outbreaks: a rapid review of the evidence. PLoS One 2020 Oct 29;15(10):e0241473 [FREE Full text] [doi: 10.1371/journal.pone.0241473] [Medline: 33120402]
14. Zimmermann BM, Fiske A, Prainsack B, Hangel N, McLennan S, Buyx A. Early perceptions of COVID-19 contact tracing apps in German-speaking countries: comparative mixed methods study. J Med Internet Res 2021 Feb 08;23(2):e25525 [FREE Full text] [doi: 10.2196/25525] [Medline: 33503000]
15. Guetterman TC, Chang T, DeJonckheere M, Basu T, Scruggs E, Vydiswaran VV. Augmenting qualitative text analysis with natural language processing: methodological study. J Med Internet Res 2018 Jun 29;20(6):e231 [FREE Full text] [doi: 10.2196/jmir.9702] [Medline: 29959110]
16. Imran AS, Daudpota SM, Kastrati Z, Batra R. Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets. IEEE Access 2020;8:181074-181090. [doi: 10.1109/access.2020.3027350]
17. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global sentiments surrounding the COVID-19 pandemic on Twitter: analysis of Twitter trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]
18. Mohamed Ridhwan K, Hargreaves C. Leveraging Twitter data to understand public sentiment for the COVID‐19 outbreak in Singapore. IJIM Data Insights 2021 Nov;1(2):100021 [FREE Full text] [doi: 10.1016/j.jjimei.2021.100021]
19. Chandra R, Krishna A. COVID-19 sentiment analysis via deep learning during the rise of novel cases. PLoS One 2021 Aug 19;16(8):e0255615 [FREE Full text] [doi: 10.1371/journal.pone.0255615] [Medline: 34411112]

20.   Rekanar K, O'Keeffe IR, Buckley S, Abbas M, Beecham S, Chochlov M, et al. Sentiment analysis of user feedback on the HSE's Covid-19 contact tracing app. Ir J Med Sci 2022 Feb 18;191(1):103-112 [FREE Full text] [doi: 10.1007/s11845-021-02529-y] [Medline: 33604836]

21.   Díaz M, Johnson I, Lazar A, Piper A, Gergle D. Addressing age-related bias in sentiment analysis. 2018 Presented at: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, QC, Canada. [doi: 10.1145/3173574.3173986]

22.   Hargittai E. Potential Biases in Big Data: omitted voices on social media. Soc Sci Comput Rev 2018 Jul 30;38(1):10-24. [doi: 10.1177/0894439318788322]

23.   Hargittai E. Is bigger always better? Potential biases of big data derived from social network sites. Ann Am Acad Political Soc Sci 2015 Apr 09;659(1):63-76. [doi: 10.1177/0002716215570866]

24.   Goh T. Singapore's Phase 3 Will Go Ahead if 3 Conditions Met, Including Having 70% TraceTogether Take-up Rate. URL: https://www.straitstimes.com/singapore/health/phase-3-will-go-ahead-if-3-conditions-are-met-including-achieving-70-per-cent [accessed 2022-02-24]

25.   Tham Y. Police Can Access TraceTogether Data Only through Person Involved in Criminal Probe: Vivian Balakrishnan. URL: https://www.straitstimes.com/singapore/politics/ police-can-access-tracetogether-data-only-through-person-involved-in-criminal [accessed 2022-02-24]

26.   Syuzhet. URL: https://github.com/mjockers/syuzhet [accessed 2022-02-24]

27.   Nielsen F. A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs. URL: http://ceur-ws.org/ Vol-718/paper_16.pdf [accessed 2022-02-24]

28.   Sim D, Lim K. Coronavirus: why aren't Singapore residents using the TraceTogether contact-tracing app. South China Morning Post. 2020. URL: https://www.scmp.com/week-asia/people/article/3084903/ coronavirus-why-arent-singapore-residents-using-tracetogether [accessed 2022-02-24]

29.   Dave A, Bharti S, Patel S, Mishra S. A real-time sentiments analysis system using Twitter data. In: Intelligent and Cloud Computing. Singapore: Springer; 2021:251-258.

30.   Woo J. Policy capacity and Singapore's response to the COVID-19 pandemic. Policy Soc 2020 Sep;39(3):345-362 [FREE Full text] [doi: 10.1080/14494035.2020.1783789] [Medline: 35039725]

31.   Schaefer GO, Ballantyne A. Ethics of digital contact tracing wearables. J Med Ethics 2021 May 14:1-5 [FREE Full text] [doi: 10.1136/medethics-2020-106958] [Medline: 33990428]

32.   Ceron A, Negri F. Public policy and social media: how sentiment analysis can support policy-makers across the policy cycle. Rivista Italiana di Politiche Pubbliche 2015;10(3):309-338.

Abbreviations

CPC: Criminal Procedure Code
DCT: digital contact tracing
NRIC: National Registration Identity Card
PDPA: Personal Data Protection Act

Original Paper

Motivations Toward Using Digital Health and Exploring the Possibility of Using Digital Health for Mental Health in Bangladesh University Students: Cross-sectional Questionnaire Study

Munjireen S Sifat[1], BSc, MPH, PhD; Sandra L Saperstein[2], PhD; Naima Tasnim[3], BA; Kerry M Green[2], PhD

[1]TSET Health Promotion Research Center, Stephenson Cancer Center, University of Oklahoma, Oklahoma City, OK, United States

[2]Department of Behavioral and Community Health, School of Public Health, University of Maryland, College Park, MD, United States

[3]BRAC University, Dhaka, Bangladesh

Corresponding Author:
Munjireen S Sifat, BSc, MPH, PhD
TSET Health Promotion Research Center
Stephenson Cancer Center
University of Oklahoma
655 Research Parkway, Suite 400
Oklahoma City, OK, 73104
United States
Phone: 1 4052716872 ext 50477
Email: Munjireen-sifat@ouhsc.edu

Abstract

Background: Digital health is efficacious for the management and prevention of mental health (MH) problems. It is particularly helpful for the young adult population, who appreciate the autonomy digital health provides, and in low-income countries, where the prevalence of MH problems is high but the supply of professionals trained in MH is low.

Objective: The objectives of this study are 2-fold: to determine whether university students in Bangladesh find using digital health for MH promotion acceptable and to examine motivational factors for using digital health for MH.

Methods: This study used a cross-sectional survey to examine the likelihood that university students in Bangladesh (n=311) would use different forms of digital health platforms for MH promotion and assessed drivers of intention to use and actual use of digital health generally and digital health for MH through the lens of the Technology Acceptance Model. The results provided evidence that the university student population in Bangladesh is likely to use digital health to promote their MH.

Results: Social influence (adjusted odds ratio [aOR] 1.68, 95% CI 1.40-2.01; $P<.001$), ease of use (aOR 1.85, 95% CI 1.35-2.53; $P<.001$), and perceived usefulness (aOR 4.12, 95% CI 1.79-9.51; $P=.001$) of digital health were found to be significant drivers of the intention to use general digital health, and having an intention to use digital health (aOR 2.10, 95% CI 1.17-3.78; $P=.01$) had the greatest influence on actual use of digital health. Social influence (aOR 1.71, 95% CI 1.43-2.04; $P<.001$), perceived usefulness (aOR 8.92, 95% CI 4.18-19.04; $P<.001$), and use of general digital health (aOR 2.16, 95% CI 1.18-3.97; $P=.01$) were associated with higher intention to use digital health for MH. The use of general digital health (aOR 4.19, 95% CI 2.37-7.41; $P<.001$) was associated with the actual use of digital health for MH, as were greater non–stigma-related barriers to using traditional clinical MH services (aOR 2.05, 95% CI 1.10-3.80; $P=.02$).

Conclusions: Overall, we see that the use of digital health for MH is acceptable in this population and can be helpful for students who perceive barriers to receiving traditional care. We also gain insight into how to promote the intention to use digital health, which in turn promotes the actual use of digital health.

KEYWORDS

Introduction

Background

Digital health interventions have become widespread to fulfill the need for mental health (MH) services that are low in supply and high in demand [1]. Mobile health (mHealth) is a category within digital health defined as the use of mobile computing and communication technologies in health care and public health [1]. The most common uses of mHealth are apps for monitoring and treating chronic conditions as well as in prevention efforts [2]. mHealth interventions have been found to be beneficial for smoking cessation, adherence to care, health behavior changes, disease management, increasing physical activity [3], and attendance rates of care [4,5]. Marcolino et al [2] examined 23 systematic reviews encompassing >10,000 articles published from 2009 to 2016 and concluded that there is strong evidence to suggest that mHealth is effective in disease management, symptom improvement, and increasing the quality of life of populations.

MH is another domain in which the use of apps has shown promising results. Apps are defined as discrete and independent software that runs on a mobile device [6,7]. Mobile apps have more benefits than SMS text messaging as they can be more deeply personalized [8], visually engage the user, track progress, and be self-paced [9,10]. These features make apps an invaluable platform for the dissemination of interventions. A systematic review evaluated 5646 abstracts published between 2008 and 2013 and found 8 papers describing 5 apps targeting depression, anxiety, and substance abuse that met their inclusion criteria [11]. The review only included evidence-based MH apps that could be downloaded from app stores. The results showed significant reductions in depression, stress, and substance use [11]. Other meta-analyses support that psychological intervention content delivered through a web or mobile app can be as efficacious as face-to-face treatment for depression [12-14].

Digital health can increase the likelihood that health interventions will be delivered to otherwise hard-to-reach populations, particularly in low- and middle-income settings [2]. A systematic review assessed 6 interventions that were specific to low- and middle-income countries and found that 5 out of 6 showed benefits to participants [15]. Other advantages of digital health are convenience, ease, cost-effectiveness, scalability, personalization, and "the ability to send time-sensitive messages with an 'always on' device" [16]. Furthermore, it can reach populations who would otherwise not engage with traditional health services [17].

There are particular benefits for governments of low-income countries that need additional support for patient management [18] because digital health is potentially highly accessible in low-income countries, with 60% of low-income populations having access to a mobile phone [19]. Furthermore, internet and smartphone use are rising worldwide in high- and low-income countries alike [20]. Bangladesh is one such low-income country that has shown positive results in the use of digital health for promoting health care in Bangladesh for various health-related issues [21,22]. The Bangladesh government fosters digital development, and the United Nations recognized its efforts toward building a digital health infrastructure in 2011 [23]; as of the beginning of 2020, >99 million people use the internet in Bangladesh [24], and most of them own smartphones [25]. Although the focus of most interventions in Bangladesh has been on the use of SMS text messaging and landlines [21,22], a handful of studies have examined using apps on smartphones for health [22]. In Bangladesh, apps have been used to link village physicians to formal physicians [26] and for diabetes management [27], nutrition services [28], and maternal and child health [29]. A systematic review examined all health-related apps in Bangladesh (N=234), and a total of nine categories of apps were mentioned in the report: general health information apps, physician information apps, institutional apps, fitness apps, mother and child apps, disease-specific care apps, herbal apps, and food and nutrition apps [20]. As such, we see a large number of mobile phone apps being used for health promotion in Bangladesh; yet, none are focused on MH promotion.

Although apps show promise in Bangladesh for other health outcomes, there is a lack of literature examining the use of mHealth for MH in this population or rates of mHealth use in general. This is particularly important given the high rates of MH problems in the population and the current lack of infrastructure in Bangladesh to deal with these problems [30]. According to the World Health Organization, there is <1 (0.001%) psychiatrist available for every 100,000 people in Bangladesh [31,32]. Although there is no national surveillance system that indicates a nationally representative prevalence rate of MH disorders in Bangladesh, a systematic review estimated the prevalence of MH disorders to be between 6.5% and 31% among adults [33]. Another systematic review examining rates of suicide estimated the rate to be 39.6 per 100,000, which is triple the global rate (10.7 per 100,000) [30].

The onset of depression typically occurs from adolescence to early adulthood [34,35]. In particular, early adulthood is deemed the "most vulnerable time" for the onset of depressive symptoms in Bangladesh [30]. This time frame, along with the multiple stressors (academic pressure and new social and physical environments) that college students face, makes the university student population particularly prone to depressive symptoms [36]. Recent studies examined MH outcomes in Bangladesh university students and found high rates of depression, ranging from 47.5% [37] to 69.5% [25]. Evidence supports that, the earlier one can manage stress and depressive symptoms, the better the overall health outcomes they will have [38]. MH apps are particularly well-suited for young adults seeking help for their symptoms because this population reports a high need for autonomy [39,40]. Young adults prefer using self-help materials if they are familiar with the medium that delivers them, such as smartphones [41].

Although the rates of MH problems may be high, there is low MH literacy [33] and high stigma surrounding the topic [28]. Hossain et al [33] found that there was low awareness of MH disorders and that attitudes toward seeking help for MH were negative. They found that even those who had an MH disorder did not prioritize MH care. This is not uncommon in low-income countries in Asia [31,42], where it is believed that MH problems

are caused by religious or cultural abnormalities [43], which in turn is associated with low use of clinical services. As mHealth has been used successfully in Bangladesh for chronic disease management [20], it is possible that it can also be used to improve MH. At a minimum, the acceptability of using digital health for MH should be determined. Developing MH messaging for in-app delivery for college students in Bangladesh has the potential to reduce, manage, and prevent depression symptomatology.

The Technology Acceptance Model (TAM) is an information technology framework for understanding users' adoption and use of emerging technologies [44]. The model proposes that a user's perception of the usefulness (ie, perceived benefits) and ease of use lead to their intent to use the technology and that intention is directly related to actual use. The TAM also posits that perceptions of usefulness and ease of use are influenced by external factors such as social influences [44]. This study uses this framework to assess where a Bangladeshi population falls on the scale of accepting digital health for MH and describing their current digital health use.

Objectives

This paper aims to (1) describe the likelihood that students will use different forms of digital health platforms for MH promotion; (2) assess the relationship between the perceived ease of use, usefulness, and social influence on the use of digital health and the intention to use and actual use of digital health; and (3) assess the relationship between the perceived ease of use, usefulness, and social influence on the use of digital health for MH and the intention to use and actual use of digital health for MH.

Methods

Study Sample

Adult university students across Bangladesh were invited to take an anonymous web-based survey. Students were emailed a flyer invitation by faculty to participate in the study and offered a 1-in-4 chance to win 422 Bangladeshi taka (US $5) for participating in the survey. In addition to faculty recruitment, flyers were posted on university social media pages. The total sample size was 311 complete responses.

Survey Creation

A total of 5 cognitive interviews were conducted with Bangladeshi university students to develop the survey. As part of the creation of the survey instruments, first, native Bangla speakers reviewed and translated the English survey items (most items were part of previously validated scales, which is explained further in the *Measures* section of this paper) into Bangla. Items with complex translations or items with cultural meanings that differed in Bangla were noted and compiled into a guide for the cognitive interviews. The cognitive interviews asked the participants to explain how they defined MH and their interpretation of the survey items. Items that were culturally inappropriate or that students did not understand were adapted to make it easier for them to understand. For example, the phrase *feeling down* to denote feeling sad or depressed in the question *How often in the past two weeks did you feel down, depressed,*

or hopeless? is not used in Bangladesh and was removed from the question. On the basis of the cognitive interviews, the questionnaire was adapted and pilot-tested with 10 participants. The pilot test participants reported no difficulties with the items, and the survey was completed.

Measures

Dependent Variables

Overview

Intention to use general digital health (Cronbach α=.88) was assessed by creating a mean score of three items: (1) *I intend to use a digital health service in the future*, (2) *I will always try to use digital health services in my daily life*, and (3) *I plan to continue to use digital health services frequently* [45]. Intention to use digital health for MH (Cronbach α=.89) was assessed similarly using the mean of three items: (1) *I intend to use digital mental health services in the future*, (2) *I will always try to use digital mental healthcare in my daily life*, and (3) *I plan to continue to use digital mental health services frequently*. Items were scored from 1 (do not agree) to 7 (totally agree) and then dichotomized into no or low (1-4.44) and moderate or high (4.45-7) intention. The cutoff was 4.45 because 4 was considered *neither agree nor disagree* and 5 was considered *slightly agree* on the scale.

Current Use

The use of digital health for general health was assessed by asking if the following statement—*I use digital health services to better my health (excluding use for mental health) currently*—was true or false. An example was given in the question stem *For example, using an app to track steps, for weight loss, to increase physical activity*. Similarly, digital health for MH was a binary variable as to whether participants used digital health: *I use digital health for mental health currently (for example, following meditation videos)*. These questions have been used in previous studies assessing use of digital health for MH in populations in low-income countries [46].

Independent Variables

Barriers to Using Clinical MH Services

Barriers to seeking MH services were measured using the Barriers to Access to Care Evaluation scale [47]. The scale consists of both stigma-related and nonstigma-related items. The participants responded using a Likert-scale of 1=not at all (indicating this was not a barrier to care) to 4=a lot (indicating a great barrier to care) to the following question: *Have any of these issues ever stopped, delayed or discouraged you from getting, or continuing with, professional care for a mental health problem?* The respondents rated how much of a barrier the provided scenarios were to receiving MH care, example barriers being *Thinking that professional care probably would not help* or *Concern about what people at work might think, say, or do*. Responses to the items were averaged to create the final score (1-4). The nonstigma (attitudinal and instrumental) barriers subscale of the Barriers to Access to Care Evaluation scale included 22 items and had adequate reliability, with a Cronbach

α of .76. The stigma subscale consisted of 12 items and had a Cronbach α of .89.

Variables in relation to general digital health and digital health specifically for MH promotion were included.

Likelihood of Using Digital Health

This variable was assessed by asking how likely the participants were to (1) text a helpline or crisis center, (2) text a professional (ie, therapist or physician), (3) use a smartphone app for self-paced meditation or nonclinical practices, (4) use a smartphone app to look up information and symptoms about MH, (5) use internet-based self-paced programs for meditation or nonclinical practices, and (6) use internet-based programs to video chat with a professional on a scale of 1 (extremely unlikely) to 5 (extremely likely).

Ease of Use of Digital Health

This variable was assessed by averaging the scores of 6 items measured on a 7-point scale (1=do not agree, 7=strongly agree). Example items included *Learning how to use digital health services is easy for me*, *My interaction with digital health service is clear and understandable*, and *I find digital health services easy to use* [45]. The Cronbach α for these items was .89.

Social Influence on Digital Health Use

This variable was assessed using the mean of the following three items on a 7-point scale (1=do not agree, 7=strongly agree): (1) *People who are important to me think that I should use a digital health service*, (2) *People who influence my behavior think that I should use a digital health service*, and (3) *People whose opinions that I value prefer that I use digital health service* [45]. The Cronbach α for these items was .94. Social influence regarding the use of digital health for MH was assessed using the mean of 3 items measured on a scale of 1 (do not agree) to 7 (strongly agree); for example, *People who are important to me think that I should use digital mental health services*. The Cronbach α for these items was .95.

Perceived Usefulness of General Digital Health

This variable was assessed by taking the mean of two items: *I find digital health services useful in my daily life* and *Using digital health services helps me accomplish things more quickly* [45]. Both were measured on a scale of 1 (do not agree) to 7 (totally agree). Perceived usefulness of digital health for MH was assessed by taking the mean of 3 items and dichotomizing the measure into a scale of 0 (low perceived usefulness; *do not agree* to *neither agree nor disagree*) to 1 (high perceived usefulness; *slightly agree* to *totally agree*). The construct was dichotomized as a method of addressing collinearity between this construct and the ease of use construct. The following is an example item: *I find that digital mental health services are or could be useful in my daily life*. The Cronbach α for these items was .88.

Covariates

Wellness

This variable was measured using the 5-item HERO Wellness Scale by Yaklin et al [48]. The scale assesses happiness, enthusiasm, resilience, and optimism, and had high-reliability

scores in the study sample (Cronbach α=.87). An example item—*On average, during the last seven days, how optimistic have you felt?*—was scored on a scale of 0 (not at all) to 10 (extremely). Final scores were created by summing the answers to all items and ranged from 0 to 50, with higher scores indicating higher wellness.

Perceived Stress

This variable was measured using the 4-item Perceived Stress Scale [49] and had an acceptable reliability score in the study sample (Cronbach α=.70). Questions such as *In the last month, how often have you felt that you were unable to control the important things in your life?* were answered on a scale of 0 (never) to 4 (very often) and were summed, with final scores ranging from 0 to 16 and higher scores indicating higher stress.

Depression

This variable was assessed using the 2-item ($r=0.53$; $P<.001$) Patient Health Questionnaire [50]. Questions such as *In the past two weeks, how often have you felt depressed or hopeless?* were answered on a scale of 0 (never) to 3 (almost every day). Scores were dichotomized into whether one was likely to have a major depressive disorder based on a cutoff point of 3 from the sum of the scale items.

Lifetime Suicidal Ideation

This was assessed as a binary variable (yes or no) as to whether they had ever had thoughts that they would rather be dead.

Physical Health

This variable was assessed using one item: *How would you rate your overall health?* (1=poor, 5=excellent).

Demographics

Socioeconomic status (SES) while growing up was assessed by asking *How often did your family have enough money to make ends meet?* Respondents answered on a scale of 0 (never) to 5 (always), and the answers were dichotomized into low versus high SES. Gender was measured in three categories: male, female, and gender minority. Age was measured as a continuous variable. Relationship status was assessed categorically; the participants selected if they were single, partnered (in a relationship or married), or other. Semester or year in school was categorized as first to third or first year, fourth to sixth or second year, seventh to ninth or third year, 10th to 12th or fourth year, and 13th or fourth year or higher. The degree of study was dichotomized as pursuing either a bachelor's or master's degree. Geographic location was assessed by asking if the participants lived in a rural or urban setting.

Specific digital health indicators of interest were payment methods, where responses indicated if the participants had monthly plans, pay as you go, or something else. The participants were also asked about their language preference for digital health and if they preferred their native language (Bangla), English, or something else.

Analysis

Analyses were conducted using complete case analysis with a final sample size of 311. Means, SDs, and frequencies were used to describe the data. Group differences between the primary

outcome of interest—use of digital health for MH—and demographic variables were assessed using analysis of variance and chi-square tests. Logistic regression analysis was used to examine the unadjusted relationships between individual predictors and outcomes of interest. If the unadjusted association was found to be associated at $P \le .20$, the variable was included in a final, adjusted logistic regression model. Models were shown to predict the intention to use and actual use of general digital health and digital health for MH. The models predicting actual use included hierarchical regression, with the first step showing unadjusted associations, the second step showing the model without including intention to use, and the final step (step 3) including the intention to use. Model fit statistics were reported.

Ethics Approval

The study was approved by the University of Maryland College Park (UMCP) Institutional Review Board (IRB number: 1656046-3)

Results

Overview

Descriptive statistics were used to describe the sample demographics in Table 1. Differences between those who used the primary outcome of digital MH and those who did not were examined within demographic variables. The sample was predominantly male (184/311, 59.2%), identified as heterosexual (276/311, 93.9%), not in a relationship (239/311, 76.8%), and sought a bachelor's degree (258/311, 83%). Growing up, the participants were mostly from families with a high SES (223/311, 71.7%) and from urban areas (167/311, 53.7%). The only significant differences among the variables of interest between those who used digital health for MH and those who did not were gender and whether they used general digital health. The participants who reported using digital health for MH (82/311, 26.4%) used general digital health at nearly twice the rate (57/82, 70%) of those who did not use digital health for MH (25/82, 30%), a significant difference ($P<.001$). Men were less likely to use digital health for MH than women and gender minorities—of those who did not use digital health for MH, 62.4% (143/229) were men and 37.6% (86/229) were women or gender minorities ($P=.049$).

Students had moderate levels of wellness (mean 26.58, SD 9.94), self-reported health status (mean 2.69, SD 0.87), and perceived stress (mean 8.46, SD 0.87). Approximately 43.4% (135/311) of the sample were likely to have depression, and 28% (78/311) reported lifetime suicidal ideation. Most students used a monthly plan to pay for their phones (223/311, 71.7%), owned their phones (308/311, 99%), and used a smartphone (310/311, 99.7%). In the sample, 43.4% (135/311) reported using digital health for general health, and 26.4% (82/311) used digital health for MH. Although half of the sample (115/311, 49.8%) did not have a preference between their native language (Bangla) and English, 31.5% (98/311) preferred Bangla and 18.6% (58/311) preferred English.

Respondents reported their likelihood of using different forms of digital health for MH promotion (Figure 1). Overall, a large percentage (227/302, 75.3% to 246/297, 82.9%) of the sample reported likelihood of using apps and internet-based programs. Most respondents said they would be likely to text a helpline or crisis center (170/290, 58.8%) or a professional (ie, therapist or physician; 220/292, 75.3%), use an app on a smartphone for self-paced meditation or nonclinical practices (227/302, 75.3%), look up information and symptoms about MH (229/301, 76.2%), use internet-based self-paced programs for meditation or nonclinical practices (242/302, 80.2%), or talk with a professional (250/301, 82.9%).

A correlation matrix of the independent variables used in all models is shown in Table 2. When examining the main constructs of the TAM related to general digital health, we found significant correlations between social influence and ease of use of general digital health ($r=0.316$; $P<.001$) and perceived usefulness ($r=0.241$; $P<.001$). For the variables related to digital health for MH, there were significant correlations among the ease of use of digital health construct, social influence ($r=0.256$; $P<.001$), and perceived usefulness ($r=0.366$; $P<.001$). There were also significant correlations among the control variables of interest—wellness was negatively correlated with stress ($r=-0.560$; $P<.001$) and depression ($r=0.338$; $P<.001$). Geography and SES were correlated ($r=0.190$; $P<.001$) in that those who lived in urban areas had higher SES. Perceived general health was positively correlated with wellness ($r=0.382$; $P<.001$) and negatively correlated with stress ($r=-0.292$; $P<.001$). Perceived stigma as a barrier to MH care and instrumental and attitudinal barriers to care were highly correlated ($r=0.765$; $P<.001$).

Table 1. Participant demographics (N=311).

Demographics	Overall	Did not use digital health for MH[a] (n=229)	Used digital health for MH (n=82)	Chi-square *P* value or ANOVA[b] *P* value
Age (18-41 years), mean (SD)	22.7 (1.86)	22.8 (1.74)	22.6 (2.18)	.59
Gender, n (%)				.049
Male	184 (59.2)	143 (62.4)	41 (50)	
Female and gender minority	127 (40.8)	86 (37.6)	41 (50)	
Sexual orientation, n (%)				.69
Heterosexual or straight	276 (93.9)	203 (93.5)	73 (94.8)	
Sexual minority (LGBTQA+[c])	18 (6.1)	14 (6.5)	4 (5.2)	
Childhood SES[d] , n (%)				.73
Low	88 (28.3)	66 (28.8)	22 (26.8)	
High	223 (71.7)	163 (71.2)	60 (73.2)	
Relationship status, n (%)				.65[e]
Single	239 (76.8)	179 (78.2)	60 (73.2)	
Partnered (relationship or married)	69 (22.2)	48 (21.0)	21 (25.6)	
Other (self-described)	3 (1.0)	2 (0.9)	1 (1.2)	
Semester or year in school, n (%)				.09
First to third or first year	64 (20.6)	51 (22.4)	13 (15.9)	
Fourth to sixth or second year	60 (19.4)	38 (16.7)	22 (26.8)	
Seventh to ninth or third year	62 (20.0)	43 (18.9)	19 (23.3)	
10th to 12th or fourth year	60 (19.4)	43 (18.9)	19 (23.2)	
13th or fourth year or higher	64 (20.6)	53 (23.3)	11 (13.4)	
Degree of study, n (%)				.17
Bachelor's (BS[f] or BA[g])	258 (83.0)	186 (81.2)	72 (87.8)	
Master's (MPH[h] or MBA[i])	53 (17.0)	43 (18.8)	10 (12.2)	
Geographic location, n (%)				.30
Rural	144 (46.3)	102 (44.5)	42 (51.2)	
Urban	167 (53.7)	127 (55.5)	40 (48.8)	
Wellness (0-50), mean (SD)	26.6 (9.94)	26.2 (9.73)	27.5 (10.48)	.32
Perceived stress (0-16), mean (SD)	8.46 (3.42)	8.42 (3.48)	8.53 (3.24)	.81
High depressive symptoms (>3), n (%)	135 (43.4)	104 (45.4)	31 (37.8)	.23
Suicidal ideation (lifetime), n (%)	78 (28.0)	58 (28.4)	20 (26.7)	.77
Rating of health status (1-5), mean (SD)	2.69 (0.87)	2.69 (0.89)	2.68 (0.86)	.92
Mobile phone plan, n (%)				.98
Monthly plan	223 (71.7)	164 (71.6)	59 (72)	
Pay as you go	52 (16.7)	38 (16.6)	14 (17.1)	
Other	36 (11.6)	27 (11.8)	9 (11)	
Phone ownership, n (%)				.78
Personal phone	308 (99.0)	227 (99.1)	81 (98.8)	
Shared phone	3 (1.0)	2 (0.9)	1 (1.2)	
Type of phone, n (%)				.55
Phone with internet capability	310 (99.7)	228 (99.6)	82 (100)	

Demographics	Overall	Did not use digital health for MH[a] (n=229)	Used digital health for MH (n=82)	Chi-square *P* value or ANOVA[b]*P* value
Phone without internet capability	1 (0.3)	1 (0.4)	0 (0)	
General digital health use, n (%)	135 (43.4)	78 (34.1)	57 (69.5)	<.001
Use of digital health for MH, n (%)	(82) 26.4	N/A[j]	N/A	N/A
Language preference for digital health, n (%)				.36
Bangla	98 (31.5)	77 (33.6)	21 (25.6)	
English	58 (18.6)	40 (17.5)	18 (22.5)	
Bangla or English	115 (49.8)	112 (48.9)	43 (52.4)	

[a]MH: mental health.

[b]ANOVA: analysis of variance.

[c]LGBTQA+: lesbian, gay, bisexual, transgender, queer, asexual plus other identities.

[d]SES: socioeconomic status. Item asked How often did your family have enough money to make ends meet growing up? Low=never, rarely, sometimes; high=most of the time, always.

[e]This chi-square test is not valid as n<5 for some cells.

[f]BS: bachelor of science.

[g]BA: bachelor of arts.

[h]MPH: master of public health.

[i]MBA: master of business administration.

[j]N/A: not applicable.

Figure 1. Distribution of likelihood of using digital health forms for mental health promotion (%; N=311). Somewhat likely and extremely likely were combined in the likely category. Somewhat unlikely and extremely unlikely were combined in the unlikely category.
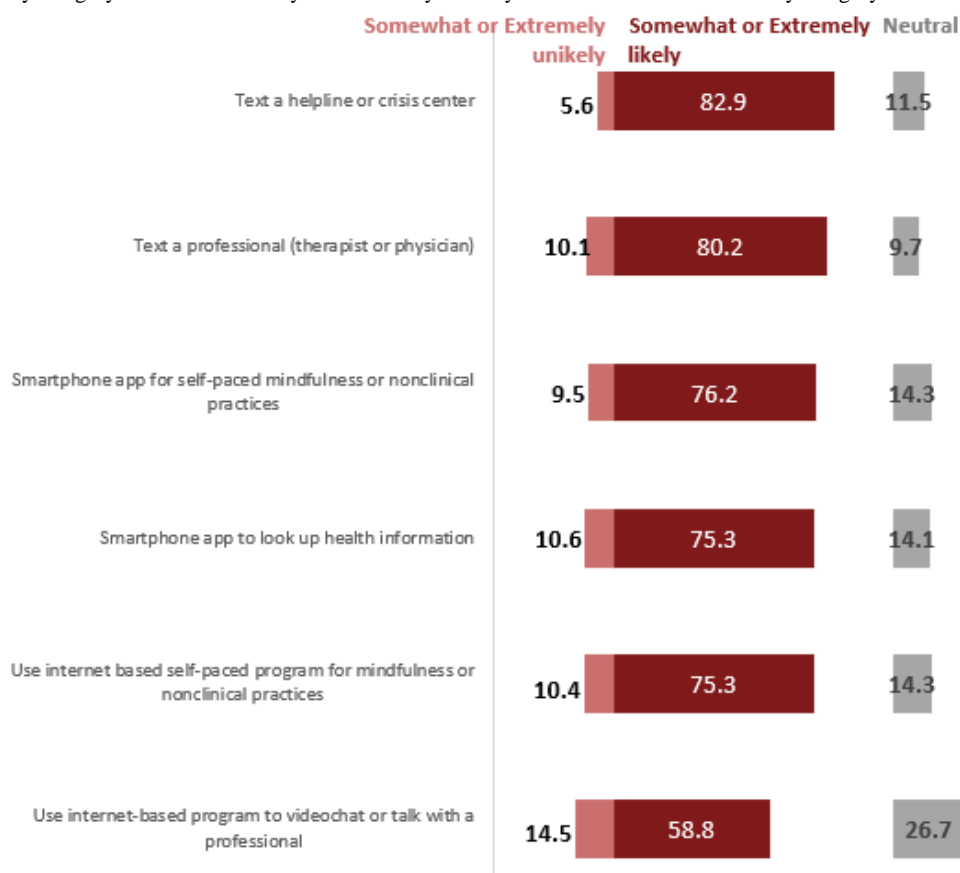
Table 2. Correlation matrix (N=311)[a].

	1[b]	2[c]	3[d]	4[e]	5[f]	6[g]	7[h]	8[i]	9[j]	10[k]	11[l]	12[m]	13[n]	14[o]
1	—[p]	.32[q]	.24[q]	.80[q]	.30[q]	−.05	.10	−.06	−.19[q]	.05	−.10	−.06	.03	.01
2	—	—	.37[q]	.26[q]	.37[q]	.10	.14[r]	−.11	−.05	.16[q]	.11	−.13[r]	−.21[q]	−.22[q]
3	—	—	—	.19[q]	.44[q]	.02	.08	−.06	−.09	.02	−.04	.01	−.14[r]	−.11
4	—	—	—	—	.33[q]	−.03	.07	−.04	−.21[q]	.01	−.06	−.10	.09	.05
5	—	—	—	—	—	.10	.04	−.01	−.10	−.05	.02	.05	−.02	.01
6	—	—	—	—	—	—	.38[q]	−.29[q]	−.17[q]	.06	.15[r]	−.22[q]	−.19[q]	−.18[q]
7	—	—	—	—	—	—	—	−.56[q]	−.19[q]	−.02	.08	−.34[q]	−.25[q]	−.25[q]
8	—	—	—	—	—	—	—	—	.17[q]	.01	.00	.45[q]	.26[q]	.24[q]
9	—	—	—	—	—	—	—	—	—	.14[r]	.10	−.03	.03	.06
10	—	—	—	—	—	—	—	—	—	—	.19[q]	−.10	−.12[r]	−.09
11	—	—	—	—	—	—	—	—	—	—	—	−.06	−.17[q]	−.20[q]
12	—	—	—	—	—	—	—	—	—	—	—	—	.25[q]	.26[q]
13	—	—	—	—	—	—	—	—	—	—	—	—	—	.76[q]
14	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Value, mean (SD)	4.15 (1.84)	5.35 (1.16)	5.37 (1.20)	4.10 (1.82)	5.15 (1.18)	2.69 (0.88)	26.6 (9.94)	8.46 (3.42)	0.41 (0.49)	1.54 (0.50)	0.72 (0.45)	0.43 (0.50)	0.70 (0.64)	0.87 (0.46)

[a]Higher scores equal greater amounts for all variables.

[b]Social influence on the use of general digital health; range: 1-7.

[c]Ease of use of general digital health; range: *poor* to *excellent*.

[d]Perceived usefulness of general digital health (0=low, 1=high).

[e]Social influence on the use of digital health for mental health (0=low, 1=high); range: 1-7.

[f]Perceived usefulness of digital health for mental health (0=low, 1=high).

[g]General health rating; range:0-50.

[h]Wellness; range: 0-50.

[i]Perceived stress; range: 0-16.

[j]Gender (0=male, 1=female).

[k]Geography (0=rural, 1=urban).

[l]Socioeconomic status (0=low, 1=high).

[m]Depression (0=low, 1=high).

[n]Stigma-related barriers to care; range: 1-4.

[o]Attitudinal and instrumental barriers to care.

[p]Not applicable.

[q]$P<.001$.

[r]$P<.01$.

Regression Analysis Results

The outcomes of intention to use and actual use of digital health in general and digital health for MH were examined using logistic regression analysis. Table 3 shows the results for the outcome of intention to use general digital health. In the unadjusted results, all the main constructs of the TAM (ease of use, social influence, and perceived usefulness of digital health) were positively associated with intention to use digital health, as was perceived wellness. In the adjusted model, these constructs remained statistically significantly associated. Those who perceived digital health to be easy to use (adjusted odds ratio [aOR] 1.85; $P<.001$), had higher approval from their social networks to use digital health (aOR 1.68; $P<.001$), and perceived higher usefulness of digital health (aOR 4.12; $P=.001$) had higher adjusted odds of intending to use digital health. The R^2 for this final adjusted model was 0.445 ($P<.001$).

XSL•FO

RenderX

Table 3. Logistic regression analysis associating the Technology Acceptance Model constructs with intention (high vs low) to use digital health (N=311)[a].

Item	Unadjusted associations		Adjusted model[b,c]	
	OR[d] (95% CI)	*P* value	aOR[e] (95% CI)	*P* value
Ease of use of digital health (1-7)[f]	2.29 (1.79-2.93)	<.001	1.85 (1.35-2.53)	<.001
Social influence on digital health use (1-7)[f]	1.79 (1.54-2.09)	<.001	1.68 (1.40-2.01)	<.001
Perceived usefulness of digital health (high vs low)	9.76 (4.70-20.35)	<.001	4.12 (1.79-9.51)	.001
Controls				
Rating of general health (poor to excellent)[f]	1.23 (0.94-1.61)	.14	1.38 (0.95-2.02)	.09
Wellness (0-50)[f]	1.02 (1.00-1.05)	.04	1.00 (0.97-1.04)	.85
Perceived stress (0-16)[f]	0.97 (0.91-1.04)	.45	N/A[g]	N/A
SES[h] growing up (high vs low)	0.67 (0.39-1.13)	.13	0.56 (0.27-1.16)	.12

[a]Geography (urban vs rural) and gender (female vs male) variables were not significant in the unadjusted model, so they were not included in the adjusted model.

[b]Nagelkerke R^2=0.445.

[c]*P* value <.001

[d]OR: odds ratio.

[e]aOR: adjusted odds ratio.

[f]Higher scores equal greater amounts.

[g]N/A: not applicable.

[h]SES: socioeconomic status.

When examining the outcome of use of digital health, the only TAM construct significantly associated was social influence, as seen in Table 4. However, when the intention to use digital health construct was added to the model (step 3 of the hierarchical regression), social influence was no longer significantly associated with the use of digital health. In the unadjusted analyses shown in step 1 of Table 4, the controls of better health, wellness, and higher SES were associated with actual use; in the adjusted model, only SES remained associated. In the final step of the model, we saw that those with higher intention to use digital health had higher adjusted odds of actual use (aOR 2.10; *P*=.01). The R^2 for this final adjusted model was 0.108 (*P*<.001).

In Table 5, for the analysis looking at predictors of intention to use digital health for MH, we found that social influence, perceived usefulness, and use of general digital health were positively associated with the intention to use digital health for MH in the unadjusted model. In the adjusted model, these constructs remained significantly associated. Higher social

influence (aOR 1.73; *P*<.001), perceived usefulness (aOR 8.70; *P*<.001), and use of general digital health (aOR 2.16; *P*=.01) were associated with higher adjusted odds of intention to use digital health for MH. The R^2 for this final adjusted model was 0.492 (*P*<.001).

The results in Table 6 show that social influence, intention to use, use of general digital health, instrumental and attitudinal barriers, and gender were positively associated with using digital health for MH. There were no significant changes when comparing step 2 of the model (without intention to use) and step 3. In the final adjusted model, we see that those who used digital health for their general health had higher odds (aOR 4.19; *P*<.001) of using digital health for MH. Those who perceived higher instrumental and attitudinal barriers to receiving clinical MH care had higher adjusted odds of using digital health for MH (aOR 2.05; *P*=.02), and women had almost twice higher adjusted odds of use when compared with men (aOR 1.88; *P*=.03). These model statistics show that the model is statistically significant (*P*<.001), with an R^2 of 0.204.

Table 4. Logistic regression associating the Technology Acceptance Model constructs with use of digital health (N=311)[a].

Item	Step 1		Step 2[b]		Step 3[c]	
	OR[d] (95% CI)	P value	aOR[e] (95% CI)	P value	aOR (95% CI)	P value
Ease of use of digital health (1-7)	1.21 (0.99-1.47)	.06	1.07 (0.86-1.33)	.56	0.97 (0.77-1.23)	.81
Social influence on digital health use (1-7)	1.14 (1.01-1.30)	.03	1.15 (1.01-1.32)	.04	1.08 (0.93-1.25)	.32
Perceived usefulness of digital health (high vs low)	1.39 (0.74-2.61)	.30	N/A[f]	N/A	N/A	N/A
Intention to use digital health (high vs low)	2.29 (1.41-3.72)	.001	N/A	N/A	2.10 (1.17-3.78)	.01
Controls						
Rating of general health (poor to excellent)	1.49 (1.14-1.94)	.004	1.37 (1.03-1.85)	.03	1.34 (0.99-1.80)	.05
Wellness (0-50)	1.03 (1.01-1.05)	.01	1.02 (0.99-1.05)	.32	1.01 (0.98-1.05)	.36
Perceived stress (0-16)	0.95 (0.89-1.02)	.15	1.00 (0.92-1.09)	.94	1.00 (0.92-1.08)	.93
SES[g] growing up (high vs low)	1.72 (1.03-2.87)	.04	1.64 (0.96-2.82)	.07	1.81 (1.04-3.12)	.04

[a]Higher scores equal greater amounts. Geography (urban vs rural) and gender (female vs male) variables were not significant in the unadjusted model; therefore, they were not included in the adjusted model.

[b]Nagelkerke R^2=0.08; P=.11.

[c]Nagelkerke R^2=0.08; P<.001.

[d]OR: odds ratio.

[e]aOR: adjusted odds ratio.

[f]N/A: not applicable.

[g]SES: socioeconomic status.

Table 5. Logistic regression associating the Technology Acceptance Model constructs with intention to use digital health for mental health (N=311)[a].

Item	Unadjusted associations		Adjusted model[b,c]	
	OR[d] (95% CI)	P value	aOR[e] (95% CI)	P value
Ease of use of digital health (1-7)	1.79 (1.44-2.23)	<.001	1.39 (0.99-1.73)	.06
Social influence on the use of digital health for mental health (1-7)	1.89 (1.61-2.21)	<.001	1.71 (1.43-2.04)	<.001
Perceived usefulness of digital health for mental health (high vs low)	15.24 (7.69-30.20)	<.001	8.92 (4.18-19.04)	<.001
Use of general digital health (yes vs no)	2.33 (1.45-3.76)	.001	2.16 (1.18-3.97)	.01
Barriers to seeking traditional clinical mental health services				
Stigma-related barriers (1-4)	0.98 (0.69-1.40)	.91	N/A[f]	N/A
Instrumental or attitudinal barriers (1-4)	1.00 (0.61-1.65)	.99	N/A	N/A
Controls				
Mental health need (need help vs not)	0.99 (0.63-1.56)	.96	N/A	N/A
Wellness (0-50)	1.02 (0.99-1.04)	.19	N/A	N/A
Perceived stress (0-16)	0.98 (0.91-1.04)	.47	N/A	N/A
SES[g] growing up (high vs low)	0.74 (0.44-1.23)	.25	N/A	N/A
Urban vs rural	0.97 (0.62-1.54)	.91	N/A	N/A
Female vs male	0.67 (0.42-1.06)	.09	1.12 (0.60-2.08)	.73

[a]Higher scores equal greater amounts.

[b] Nagelkerke R^2=0.49.

[c]$P<.001$.

[d]OR: odds ratio.

[e]aOR: adjusted odds ratio.

[f]N/A: not applicable.

[g]SES: socioeconomic status.

Table 6. Logistic regression associating the Technology Acceptance Model constructs with use of digital health for mental health (N=311)[a].

Item	Step 1		Step 2[b]		Step 3[c]	
	OR[d] (95% CI)	P value	aOR[e] (95% CI)	P value	aOR (95% CI)	P value
Ease of use of digital health (1-7)	1.20 (0.95-1.51)	.12	1.12 (0.86-1.48)	.40	1.11 (0.84-1.47)	.45
Social influence on the use of digital health for mental health (1-7)	1.18 (1.02-1.37)	.02	1.14 (0.96-1.36)	.12	1.11 (0.92-1.34)	.26
Perceived usefulness of digital health for mental health (high vs low)	1.93 (1.00-3.73)	.05	1.40 (0.65-3.04)	.40	1.26 (0.55-2.89)	.59
Intention to use digital health for mental health (high vs low)	2.29 (1.31-4.01)	.004	N/A[f]	N/A	1.31 (0.62-2.77)	.49
Use of general digital health (yes vs no)	4.41 (2.56-7.60)	<.001	4.33 (2.47-7.61)	<.001	4.19 (2.37-7.41)	<.001
Barriers to seeking traditional clinical mental health services						
Stigma-related barriers (1-4)	1.23 (0.84-1.81)	.29	N/A	N/A	N/A	N/A
Instrumental or attitudinal barriers (1-4)	1.72 (1.00-2.97)	.05	2.06 (1.11-3.82)	.02	2.05 (1.10-3.80)	.02
Controls, female vs male	1.66 (1.00-2.77)	.05	1.91 (1.08-3.36)	.03	1.88 (1.07-3.23)	.03

[a]Higher scores equal greater amounts. The unadjusted models examined 5 potential additional control variables (mental health need, wellness, perceived stress, socioeconomic status, and geography), and $P<.20$ for none of them; thus, they were excluded from the adjusted models.

[b]Nagelkerke R^2=0.20; $P<.001$.

[c]Nagelkerke R^2=0.20; $P<.001$.

[d]OR: odds ratio.

[e]aOR: adjusted odds ratio.

[f]N/A: not applicable.

Discussion

Principal Findings

This study explored the acceptability of using digital health to promote MH among university students in Bangladesh. Although the MH of the sample was comparable with other Bangladeshi university student samples [37,51], with 43.4% (135/311) of the sample experiencing symptoms of depression, we cannot compare the percentage of students who use smartphones with other studies as other studies included smartphone use as an eligibility criterion in research pertaining to the digital field [52,53]. In this sample, nearly all the students owned a personal smartphone (308/311, 99.7%). We found a similar percentage of students (135/311, 43.4%) who self-reported use of digital health, defined as answering *I use digital health services to better my health (excluding use for mental health) currently* in the affirmative, to that of Waldman et al [54], who found that 45% of their Bangladeshi student sample reported looking up health-related information on the internet. To our knowledge, no other studies have examined digital health for MH promotion in Bangladesh; however, our findings show that most students would be likely to use digital health for MH. Our findings also support previous research that shows that those who are more cognizant of their health—in this case, those who already use digital health for general health—are more likely to also be attuned to their psychological health [55]. Specifically, we found that people who use digital health in general have 4 times the odds of also using digital health for MH compared with those who do not use digital health at all.

Overall, the findings partially confirm the hypotheses that the constructs of the TAM (perceived ease of use, usefulness, and social influence) are essential precursors of intention to use and actual use of digital health in general and digital health for MH. We found that the TAM constructs were particularly useful in predicting the intention to use digital health both generally and for MH, as can be seen by the model fit statistics, which included control variables (Nagelkerke R^2=0.445-0.492).

Interestingly, the TAM constructs were predictive of the intention to use both general digital health and digital health for MH but not for the actual use of either (in the adjusted models); this may be because university students in Bangladesh are unaware of or do not have the type of digital health that they would prefer to use. This aligns somewhat with the TAM as the model posits that intention mediates the connection between these constructs and actual use [44]. Therefore, theoretically, ease of use, social influence, and perceived usefulness should be stronger predictors of intention. We also found intention to be a strong predictor of actual use in the unadjusted analysis. These findings suggest that it is necessary for people to think that the use of a product (in this case, a digital health platform) is approved by their social network, easy to use, and valuable for them to form an intent to use the product and that intention is an important precursor to action. This research found that Bangladeshi students would like to use digital health for MH. However, future research should examine whether user-friendly digital health products that promote MH exist for this population to use as previous research shows that usability issues are a barrier to using digital health in general [20].

The results suggest that increasing the use of digital health in general would promote the use of digital health for MH as well. According to our results, one way to do this is to increase the acceptability of digital health products among peer groups as social influence is predictive for general and MH-specific digital health use (and intention toward use). Using user-centered design to ensure that the product is easy to use and meets the users' needs is also imperative as ease of use and perceived usefulness of an app are associated with intention to use and actual use of general digital health and the intention to use digital health for MH. From an implementation science perspective, this information is critical when developing health communication strategies around health promotion [56]. University administrations can use these findings to encourage transparency regarding health promotion directly, which may indirectly affect the use of digital health for MH.

The results show that university students are open to the use of digital health for MH in Bangladesh and that digital health may be particularly useful for those who may not otherwise seek clinical care as those with higher instrumental and attitudinal barriers are twice as likely to use digital health for MH. We also found that women had higher odds of using digital health for MH, although previous research in rural Bangladesh found that women were less aware of mHealth services than men even though intentions to use mHealth were high regardless of gender [57]. This difference in findings may be because, in this sample, university students had access to mobile phones across genders, unlike in the rural sample, where women had lower rates of phone ownership than men [57]; as such, our findings may show that, when given access, women translate their intention to use mHealth into actual use of such platforms. Previous research in Bangladesh found that women are also less likely to seek physical health care than men [58], although gender differences in MH service use have not been assessed in this population. In other populations where gender differences in MH service use have been researched, it has been found that women seek traditional MH care more than men [59]. In any case, these study results indicate that an MH promotion program on a digital platform may benefit this subgroup of people who may face attitudinal or instrumental barriers to seeking clinical care. This information can be key in marketing as the app can be framed so that it can be used autonomously. To our knowledge, no current app exists that was developed with empirical evidence for a Bangladeshi student population; as such, next steps would entail developing and pilot-testing such an app.

Limitations

The limitations of this study stem from its cross-sectional design and convenience sample. Owing to the study design, the results were not able to determine causality, nor are they generalizable to the Bangladeshi university student populations at large. As the study's topic was MH promotion, it is possible that respondents were inclined toward this approach, and those who were not interested in MH promotion did not participate. We cannot conclude that a digital MH promotion program would be of interest to all university students in Bangladesh; however, we can say that this type of program shows promise for students who may not be inclined to receive clinical MH care.

Conclusions and Future Directions

This study is the first of its kind to examine the acceptability of digital health for MH in Bangladesh. The results show that students are quite open to using digital health as a tool to improve their MH and highlight the influence that social networks might have on this decision-making. As these findings provide evidence of the acceptability of using digital health for MH, future research should pilot-test messaging for a self-paced MH app, such as a meditation app or a web-based intervention, in Bangla. Universities should promote these mental wellness programs to their students.

Authors' Contributions

MSS developed the research question, study instrument, research design, and data analysis, and wrote the manuscript. SLS consulted on data analysis and the theoretical underpinnings of the research. NT assisted with translation of the survey. KMG consulted on the research design and data analysis and edited the paper.

Conflicts of Interest

None declared.

References

1. Free C, Phillips G, Felix L, Galli L, Patel V, Edwards P. The effectiveness of M-health technologies for improving health and health services: a systematic review protocol. BMC Res Notes 2010 Oct 06;3:250 [FREE Full text] [doi: 10.1186/1756-0500-3-250] [Medline: 20925916]

2. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]

3. Fanning J, Mullen SP, McAuley E. Increasing physical activity with mobile devices: a meta-analysis. J Med Internet Res 2012;14(6):e161 [FREE Full text] [doi: 10.2196/jmir.2171] [Medline: 23171838]

4. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012;12(12):CD007459 [FREE Full text] [doi: 10.1002/14651858.CD007459.pub2] [Medline: 23235644]

5. Guy R, Hocking J, Wand H, Stott S, Ali H, Kaldor J. How effective are short message service reminders at increasing clinic attendance? A meta-analysis and systematic review. Health Serv Res 2012;47(2):614-632 [FREE Full text] [doi: 10.1111/j.1475-6773.2011.01342.x] [Medline: 22091980]

6. Heffner JL, Vilardaga R, Mercer LD, Kientz JA, Bricker JB. Feature-level analysis of a novel smartphone application for smoking cessation. Am J Drug Alcohol Abuse 2015;41(1):68-73 [FREE Full text] [doi: 10.3109/00952990.2014.977486] [Medline: 25397860]

7. Lui JH, Marcus DK, Barry CT. Evidence-based apps? A review of mental health mobile applications in a psychotherapy context. Prof Psychol Res Pr 2017;48(3):199-210. [doi: 10.1037/pro0000122]

8. Gustafson DH, McTavish FM, Chih MY, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]

9. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. Drug Alcohol Depend 2014;143:87-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.07.006] [Medline: 25085225]

10. Luxton DD, McCann RA, Bush NE, Mishkind MC, Reger GM. mHealth for mental health: integrating smartphone technology in behavioral healthcare. Prof Psychol Res Pr 2011;42(6):505-512. [doi: 10.1037/a0024485]

11. Donker T, Petrie K, Proudfoot J, Clarke J, Birch MR, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]

12. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]

13. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]

14. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol Med 2010;40(12):1943-1957. [doi: 10.1017/S0033291710000772] [Medline: 20406528]

15. Hall CS, Fottrell E, Wilkinson S, Byass P. Assessing the impact of mHealth interventions in low- and middle-income countries--what has been shown to work? Glob Health Action 2014;7:25606 [FREE Full text] [doi: 10.3402/gha.v7.25606] [Medline: 25361730]

16. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016;4(4):CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]

17. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]

18. Aranda-Jan CB, Mohutsiwa-Dibe N, Loukanova S. Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. BMC Public Health 2014;14:188 [FREE Full text] [doi: 10.1186/1471-2458-14-188] [Medline: 24555733]

19. Bastawrous A, Armstrong MJ. Mobile health use in low- and high-income countries: an overview of the peer-reviewed literature. J R Soc Med 2013;106(4):130-142 [FREE Full text] [doi: 10.1177/0141076812472620] [Medline: 23564897]

20. Islam MN, Karim MM, Inan TT, Islam AK. Investigating usability of mobile health applications in Bangladesh. BMC Med Inform Decis Mak 2020;20(1):19 [FREE Full text] [doi: 10.1186/s12911-020-1033-3] [Medline: 32013965]

21. Ahmed T, Rizvi SJ, Rasheed S, Iqbal M, Bhuiya A, Standing H, et al. Digital health and inequalities in access to health services in Bangladesh: mixed methods study. JMIR Mhealth Uhealth 2020;8(7):e16473 [FREE Full text] [doi: 10.2196/16473] [Medline: 32706736]

22. Ahmed T, Bloom G, Iqbal M, Lucas H, Rasheed S, Waldman L, et al. E-health and M-health in Bangladesh: opportunities and challenges. IDS Evid Rep 2014(60) [FREE Full text]

23. Ministry of Health and Family Welfare. Health bulletin 2012. Directorate General of Health Services. Dhaka, Bangladesh: Management Information System (MIS), Government of the People's Republic of Bangladesh; 2012. URL: https://dghs.gov.bd/ [accessed 2022-02-09]

24.   Internet subscribers in Bangladesh January, 2020. Bangladesh Telecommunication Regulatory Commission. URL: http:/ /www.btrc.gov.bd/telco/internet [accessed 2022-02-09]

25.   Islam MS, Sujan MS, Tasnim R, Mohona RA, Ferdous MZ, Kamruzzaman S, et al. Problematic smartphone and social media use among Bangladeshi college and university students amid COVID-19: the role of psychological well-being and pandemic related factors. Front Psychiatry 2021;12:647386 [FREE Full text] [doi: 10.3389/fpsyt.2021.647386] [Medline: 33935834]

26.   Khan NU, Rasheed S, Sharmin T, Ahmed T, Mahmood SS, Khatun F, et al. Experience of using mHealth to link village doctors with physicians: lessons from Chakaria, Bangladesh. BMC Med Inform Decis Mak 2015;15:62 [FREE Full text] [doi: 10.1186/s12911-015-0188-9] [Medline: 26242574]

27.   Yasmin F, Ali L, Banu B, Rasul FB, Sauerborn R, Souares A. Understanding patients' experience living with diabetes type 2 and effective disease management: a qualitative study following a mobile health intervention in Bangladesh. BMC Health Serv Res 2020;20(1):29 [FREE Full text] [doi: 10.1186/s12913-019-4811-9] [Medline: 31918704]

28.   Uddin J, Biswas T, Adhikary G, Ali W, Alam N, Palit R, et al. Impact of mobile phone-based technology to improve health, population and nutrition services in Rural Bangladesh: a study protocol. BMC Med Inform Decis Mak 2017;17(1):101 [FREE Full text] [doi: 10.1186/s12911-017-0502-9] [Medline: 28683742]

29.   Alam M, D'Este C, Banwell C, Lokuge K. The impact of mobile phone based messages on maternal and child healthcare behaviour: a retrospective cross-sectional survey in Bangladesh. BMC Health Serv Res 2017;17(1):434 [FREE Full text] [doi: 10.1186/s12913-017-2361-6] [Medline: 28645278]

30.   Arafat SM. Current challenges of suicide and future directions of management in Bangladesh: a systematic review. Global Psychiatry 2019;2(1):9-20. [doi: 10.2478/gp-2019-0001]

31.   Giasuddin NA, Levav I, Gal G. Mental health stigma and attitudes to psychiatry among Bangladeshi medical students. Int J Soc Psychiatry 2015;61(2):137-147. [doi: 10.1177/0020764014537237] [Medline: 24913265]

32.   World Health Organization. WHO-AIMS Report on Mental Health System in Bangladesh. New Delhi, India: WHO Regional Office for South-East Asia; 2007.

33.   Hossain MD, Ahmed HU, Chowdhury WA, Niessen LW, Alam DS. Mental disorders in Bangladesh: a systematic review. BMC Psychiatry 2014;14:216 [FREE Full text] [doi: 10.1186/s12888-014-0216-9] [Medline: 25073970]

34.   Kessler RC, Angermeyer M, Anthony JC, DE Graaf R, Demyttenaere K, Gasquet I, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry 2007;6(3):168-176 [FREE Full text] [Medline: 18188442]

35.   McGorry P. Transition to adulthood: the critical period for pre-emptive, disease-modifying care for schizophrenia and related disorders. Schizophr Bull 2011;37(3):524-530 [FREE Full text] [doi: 10.1093/schbul/sbr027] [Medline: 21505119]

36.   Eisenberg D, Hunt J, Speer N, Zivin K. Mental health service utilization among college students in the United States. J Nerv Ment Dis 2011;199(5):301-308. [doi: 10.1097/NMD.0b013e3182175123] [Medline: 21543948]

37.   Sayeed A, Hassan MN, Rahman MH, El Hayek S, Banna MH, Mallick T, et al. Facebook addiction associated with Internet activity, depression and behavioral factors among university students of Bangladesh: a cross-sectional study. Child Youth Serv Rev 2020;118(C):105424. [doi: 10.1016/j.childyouth.2020.105424]

38.   Bukh JD, Bock C, Vinberg M, Kessing LV. The effect of prolonged duration of untreated depression on antidepressant treatment outcome. J Affect Disord 2013;145(1):42-48. [doi: 10.1016/j.jad.2012.07.008] [Medline: 22854096]

39.   Fuller-Tyszkiewicz M, Richardson B, Klein B, Skouteris H, Christensen H, Austin D, et al. A mobile app-based intervention for depression: end-user and expert usability testing study. JMIR Ment Health 2018;5(3):e54 [FREE Full text] [doi: 10.2196/mental.9445] [Medline: 30139722]

40.   Wilson CJ, Rickwood DJ, Bushnell JA, Caputi P, Thomas SJ. The effects of need for autonomy and preference for seeking help from informal sources on emerging adults' intentions to access mental health services for common mental disorders and suicidal thoughts. Adv Ment Health 2011;10(1):29-38. [doi: 10.5172/jamh.2011.10.1.29]

41.   Martinez R, Williams C. Matching clients to CBT self-help resources. In: Bennett-Levy J, Richards D, Farrand P, Christensen H, Griffiths K, Kavanagh D, et al, editors. Oxford Guide to Low Intensity CBT Interventions. Oxford, UK: Oxford University Press; 2010:113-120.

42.   Lauber C, Rössler W. Stigma towards people with mental illness in developing countries in Asia. Int Rev Psychiatry 2007;19(2):157-178. [doi: 10.1080/09540260701278903] [Medline: 17464793]

43.   Ndetei D, Khasakhala LI, Mutiso V, Mbwayo AW. Knowledge, attitude and practice (KAP) of mental illness among staff in general medical facilities in Kenya: practice and policy implications. Afr J Psychiatry (Johannesbg) 2011;14(3):225-235. [doi: 10.4314/ajpsy.v14i3.6] [Medline: 21863208]

44.   Portz JD, Bayliss EA, Bull S, Boxer RS, Bekelman DB, Gleason K, et al. Using the technology acceptance model to explore user experience, intent to use, and use behavior of a patient portal among older adults with multiple chronic conditions: descriptive qualitative study. J Med Internet Res 2019;21(4):e11604 [FREE Full text] [doi: 10.2196/11604] [Medline: 30958272]

45.   Hoque MR, Sorwar G. Factors influencing mHealth acceptance among elderly people in Bangladesh. arXiv (forthcoming) 2016.

46. Ben-Zeev D, Fathy C, Jonathan G, Abuharb B, Brian RM, Kesbeh L, et al. mHealth for mental health in the Middle East: need, technology use, and readiness among Palestinians in the West Bank. Asian J Psychiatr 2017;27:1-4. [doi: 10.1016/j.ajp.2017.02.010] [Medline: 28558877]

47. Clement S, Brohan E, Jeffery D, Henderson C, Hatch SL, Thornicroft G. Development and psychometric properties the Barriers to Access to Care Evaluation scale (BACE) related to people with mental ill health. BMC Psychiatry 2012;12:36 [FREE Full text] [doi: 10.1186/1471-244X-12-36] [Medline: 22546012]

48. Yaklin S, Jain R, Cole SP, Raison C, Rolin D, Jain S. HERO Wellness Scale: examining a new mental wellness scale. Ann Clin Psychiatry 2020;32(1):33-40. [Medline: 31990967]

49. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983;24(4):385-396. [Medline: 6668417]

50. Ganguly S, Samanta M, Roy P, Chatterjee S, Kaplan DW, Basu B. Patient health questionnaire-9 as an effective tool for screening of depression among Indian adolescents. J Adolesc Health 2013;52(5):546-551. [doi: 10.1016/j.jadohealth.2012.09.012] [Medline: 23299020]

51. Koly KN, Abdullah R, Shammi FA, Akter T, Hasan MT, Eaton J, et al. Mental health and community-based rehabilitation: a qualitative description of the experiences and perspectives of service users and carers in Bangladesh. Community Ment Health J 2022;58(1):52-66. [doi: 10.1007/s10597-021-00790-0] [Medline: 33683536]

52. Rashid M, Hossain MB. Impact of smartphone on academic performance: a case study on the female students of a public university of Bangladesh. Int J Nat Soc Sci 2019;6(2):54-62 [FREE Full text]

53. Hossain ME, Ahmed SM. Academic use of smartphones by university students: a developing country perspective. Electron Libr 2016;34(4):651-665. [doi: 10.1108/el-07-2015-0112]

54. Waldman L, Ahmed T, Scott N, Akter S, Standing H, Rasheed S. 'We have the internet in our hands': Bangladeshi college students' use of ICTs for health information. Global Health 2018;14(1):31 [FREE Full text] [doi: 10.1186/s12992-018-0349-6] [Medline: 29554929]

55. Kruk M, Zarychta K, Horodyska K, Boberska M, Scholz U, Radtke T, et al. What comes first, negative emotions, positive emotions, or moderate-to-vigorous physical activity? Ment Health Phys Act 2019;16:38-42. [doi: 10.1016/j.mhpa.2019.03.002]

56. Cabassa LJ, Baumann AA. A two-way street: bridging implementation science and cultural adaptations of mental health treatments. Implement Sci 2013;8:90 [FREE Full text] [doi: 10.1186/1748-5908-8-90] [Medline: 23958445]

57. Khatun F, Heywood AE, Hanifi SM, Rahman MS, Ray PK, Liaw ST, et al. Gender differentials in readiness and use of mHealth services in a rural area of Bangladesh. BMC Health Serv Res 2017;17(1):573 [FREE Full text] [doi: 10.1186/s12913-017-2523-6] [Medline: 28821243]

58. Ahmed SM, Adams AM, Chowdhury M, Bhuiya A. Gender, socioeconomic development and health-seeking behaviour in Bangladesh. Soc Sci Med 2000;51(3):361-371. [doi: 10.1016/s0277-9536(99)00461-x] [Medline: 10855923]

59. Smith KL, Matheson FI, Moineddin R, Dunn JR, Lu H, Cairney J, et al. Gender differences in mental health service utilization among respondents reporting depression in a national health survey. Health 2013;5(10):1561-1571. [doi: 10.4236/health.2013.510212]

Abbreviations

aOR: adjusted odds ratio
MH: mental health
mHealth: mobile health
SES: socioeconomic status
TAM: Technology Acceptance Model

<u>Original Paper</u>

Discussions of Asperger Syndrome on Social Media: Content and Sentiment Analysis on Twitter

Elia Gabarron[1,2], PhD; Anders Dechsling[1], MSci; Ingjerd Skafle[3], MSci; Anders Nordahl-Hansen[1], Prof Dr

[1]Department of Education, ICT and Learning, Østfold University College, Halden, Norway

[2]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

[3]Faculty of Health, Welfare and Organisation, Østfold University College, Kråkerøy, Norway

Corresponding Author:
Elia Gabarron, PhD
Department of Education, ICT and Learning
Østfold University College
B R A Veien 4
Halden, 1757
Norway
Phone: 47 94863460
Email: egabarron@gmail.com

Abstract

Background: On May 8, 2021, Elon Musk, a well-recognized entrepreneur and business magnate, revealed on a popular television show that he has Asperger syndrome. Research has shown that people's perceptions of a condition are modified when influential individuals in society publicly disclose their diagnoses. It was anticipated that Musk's disclosure would contribute to discussions on the internet about the syndrome, and also to a potential change in the perception of this condition.

Objective: The objective of this study was to compare the types of information contained in popular tweets about Asperger syndrome as well as their engagement and sentiment before and after Musk's disclosure.

Methods: We extracted tweets that were published 1 week before and after Musk's disclosure that had received >30 likes and included the terms "Aspergers" or "Aspie." The content of each post was classified by 2 independent coders as to whether the information provided was valid, contained misinformation, or was neutral. Furthermore, we analyzed the engagement on these posts and the expressed sentiment by using the AFINN sentiment analysis tool.

Results: We extracted a total of 227 popular tweets (34 posted the week before Musk's announcement and 193 posted the week after). We classified 210 (92.5%) of the tweets as neutral, 13 (5.7%) tweets as informative, and 4 (1.8%) as containing misinformation. Both informative and misinformative tweets were posted after Musk's disclosure. Popular tweets posted before Musk's disclosure were significantly more engaging (received more comments, retweets, and likes) than the tweets posted the week after. We did not find a significant difference in the sentiment expressed in the tweets posted before and after the announcement.

Conclusions: The use of social media platforms by health authorities, autism associations, and other stakeholders has the potential to increase the awareness and acceptance of knowledge about autism and Asperger syndrome. When prominent figures disclose their diagnoses, the number of posts about their particular condition tends to increase and thus promote a potential opportunity for greater outreach to the general public about that condition.

KEYWORDS

XSL•FO
RenderX

Introduction

Background

Asperger syndrome (hereafter referred to as Asperger), which was removed from the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) as a formal diagnosis, is currently merged with autism and pervasive developmental disorders that are otherwise not specified within "autism spectrum disorder" (hereafter referred to as autism). Although the International Classification of Diseases framework in its current 10th edition has not yet formally removed Asperger as a diagnosis; it will do so within the 11th version that is forthcoming [1]. However, those that have been diagnosed with Asperger may choose to use this as their formal diagnosis. The removal of Asperger from the DSM-5 has been controversial as many individuals took, and continue to take, pride in their Asperger identity [2]; however, many people with Asperger were also in favor of subsuming Asperger into the broader category of autism as a spectrum of conditions [3] on the grounds of, for instance, equality of service provision and legal protection. Although still a term used by many, Asperger is becoming more and more known as an integrated part of the broader autism spectrum and thus also a part of the blooming autistic neurodiversity movement [4]. Although less focus has been given to the term Asperger in recent years, public figures who are open about their diagnosis contribute to increased media attention. Examples of highly prominent people with Asperger who have been publicly open about their diagnosis include Elon Musk (who revealed his condition on Saturday Night Live on May 8, 2021) and Greta Thunberg (who refers to her Asperger diagnosis as her "superpower"). The media attention given to such disclosures typically leads to subsequent social media discussions that can inform but also misinform public perception [5]. Asperger and autism have become a part of pop culture [6,7], and research indicates there are advantages and disadvantages to this from a public health perspective but also, more importantly, an impact on individuals with the diagnosis. Asperger and autism are misunderstood and often "hidden." They are hidden because there are no facial or bodily features that readily indicate a person has these conditions or not. Misunderstandings of the diagnoses relate directly to media representations (eg, films such as Rain Man [8] that are so influential that they can be perceived by the public as definitions for a whole diagnosis). For instance, savant syndrome is widely misconceived as very common in the population of individuals with autism and because of this and the intriguing features of the syndrome, it is a common part of media representation [9] of individuals with autism. However, the prevalence of savant syndrome in people with autism is rare [10]. Another widely held misconception of autism and Asperger is that people with autism, in general, are asocial [11]. Media attention is a double-edged sword in that public awareness and acceptance can increase, but it can also oversimplify highly complex heterogeneous conditions such as autism and Asperger. Oversimplifications and misconceptions can lead to stereotype thinking [12] and can maintain stigmas [13], but this could be reduced with the increase in societal awareness and understanding of the conditions [14]. This knowledge could be acquired or promoted through social media, where millions of people are exposed daily to all varieties of information, whether they explicitly seek it or find it unintentionally. Research on different conditions shows that health promotion through social media is mostly linked to positive effects [15-20].

Previous research has shown that the combination of key figures in society and their disclosures of personal diagnoses can affect the public's perception of their conditions [21,22]. As with Musk's disclosure and in the age of social media, the expectation was that the number of web-based discussions about Asperger and autism would increase, which could contribute further to spreading knowledge and awareness and thus potentially alter people's perceptions of the condition.

Objectives

The objective of this study was to compare the types of content used in (informative, misinformative, or neutral), the engagement with, and the sentiment of popular tweets about Asperger before and after Musk's disclosure.

Methods

Study Design and Sample

We designed a cross-sectional study to analyze the type of content, engagement, and perception of popular tweets related to Asperger. We defined popular tweets as those with more than 30 likes as this is slightly higher than the average of 25 likes reported in previous research [23]. We focused on popular tweets because of their potential to reach more users.

Eligibility Criteria for Tweets

We used the Twitter advanced search engine to find popular tweets (>30 likes) that were posted in English and included the terms "Aspergers" or "Aspie." The term "Aspie" was included because it is a common slang term used by the autism community to refer to Asperger [5]. We extracted tweets that were published between May 1, 2021, and May 14, 2021, (1 week before and after Musk disclosed his Asperger diagnosis on May 8, 2021). As commonly used on other social media platforms, the "like" function on Twitter was chosen because of the quickness and ease (one click) with which it allows users to show their agreement with a posted tweet compared to the retweet function, which requires at least 2 clicks and potentially, the addition of accompanying text from the media user [24].

Data Extraction and Classification

From the selected popular tweets, we extracted the post message and the number of comments, retweets, and likes. We extracted only the original tweets and no personal or identifiable data were collected.

We created a coding guideline, which was used to classify the tweets (the coding guideline is available in the data repository) [25]. Each post was classified by 2 independent coders who are experts in autism (IS and ANH) as to whether the post contained correct information (eg, "people with Asperger have difficulties in social interaction"), misinformation (eg, "people with Asperger have an IQ below average"), or neutral information (eg, "my son has Asperger"). Classification disagreements were

discussed with the rest of the coauthors until a consensus was reached. We analyzed the number of posts and engagement from other media users with those posts. The engagement was assessed using the number of comments, retweets, and likes.

Sentiment Analysis

We analyzed the perceptions or sentiments of each post using the AFINN sentiment analysis tool [26]. AFINN is a lexicon that assigns scores to each word ranging from –5 (very negative) to 5 (very positive) [25]. The text-mining tool considers scores higher than 0 as positive words, and scores lower than 0 as negative words [26]. The AFINN tool has been used to analyze the sentiment included in tweets about Asperger [5], and to compare the sentiment towards different COVID-19 vaccines expressed in social media posts [27]. The AFINN tool has also been used to analyze free short text, such as answers given in surveys [28,29], web-based reviews [30], self-reported notes [31], or descriptions of public health campaigns [32].

Data Analysis

All statistical analyses were performed using SPSS (version 25.0; IBM Corp). Both the data set and data analysis, including scripts, were made available in the data repository [25]. The treatment of data for this study was approved by the data

protection officer at the University Hospital of North Norway (Nr.02489).

Results

We extracted a total of 227 popular tweets. A total of 34 (14.9%) tweets were posted the week before the Musk announcement and 193 (85%) were posted the following week.

Engagement and Sentiment of Tweets Before and After Musk's Announcement

When we compared the engagement and sentiment of tweets before and after Musk's announcement, we found that tweets posted before the disclosure received significantly more engagement; they received more comments (254.15, 95% CI 87.1 to 331.5 compared to 44.88, 95% CI –74.6 to 493.2; $P<.001$), more retweets (494.47, 95% CI 28.3 to 635.6 compared to 190.80, 95% CI –321.6 to 928.9; $P=.001$), and more likes (7058.00, 95% CI 1734.9 to 10,443.6 compared to 969.44, 95% CI –4483.2 to 16,661.8; $P<.001$) than after the announcement (See Table 1). We did not find any statistically significant differences regarding the sentiment of tweets posted before and after Musk's disclosure. Finally, the Kruskal-Wallis test did not show significant differences regarding engagement or sentiment according to the type of content provided by the tweet.

Table 1. Engagement with and sentiment of popular tweets about Asperger according to the time points when they were posted and the type of information provided.

Category of engagement	Time point when the tweet was posted		t test[a]	P value	Type of information provided			P value[b]
	Before Musk's disclosure (n=34), mean (95% CI)	After Musk's disclosure (n=193), mean (95% CI)			Provides information (n=13)[c], mean (95% CI)	Neutral tweets (n=210)[d], mean (95% CI)	Contains misinformation (n=4), mean (95% CI)	
Comments	254.15 (87.08 to 331.46)	44.88 (–74.63 to 493.16)	3.375	<.001	68.77 (–43.36 to 180.93)	77.96 (30.12 to 125.79)	9.50 (–0.77 to 19.77)	.81
Retweets	494.47 (–28.30 to 635.65)	190.80 (–321.57 to 928.91)	1.803	.001	146.66 (–47.94 to 340.86)	246.00 (117.76 to 374.23)	18.25 (–7.29 to 43.79)	.59
Likes	7058.00 (1734.96 to 10,443.57)	969.44 (–4483.23 to 16,661.76)	2.756	.001	824.00 (–216.54 to 1864.54)	1980.41 (277.11 to 3683.72)	124.75 (–26.96 to 276.46)	.26
Sentiment	0.12 (–1.42 to 1.08)	0.29 (–1.29 to 0.95)	0.272	.22	–1.46 (–3.20 to 0.28)	0.38 (–0.09 to 0.84)	0.00 (–2.60 to 2.60)	.21

[a]df=225.

[b]P value obtained from the Kruskal-Wallis test.

[c]Example of a tweet that provides information: "Not all autistic people (including people with Asperger's diagnoses) are white, male techie types. Some of us are poets. Some of us are even women."

[d]Example of a neutral tweet: "Elon Musk reveals he has Asperger's syndrome during SNL monologue."

Informative, Neutral, and Misinformative Tweets Following Musk's Announcement

The interrater agreement for the classification of the tweets and respective tweet data provided was κ=0.469 (moderate agreement). We classified 210 (92.5%) of the 227 tweets as being neutral, 13 (5.7%) tweets as informative, and 4 (1.8%) as containing misinformation. Both informative and misinformative tweets were posted after Musk's disclosure. Tweets identified as misinformative included the following examples: tweets suggesting that Musk's Asperger was deeply problematic, a

tweet suggesting that autism and Asperger are the results of asymmetrical brain stem injuries, and a joke related to Musk's development of the SpaceX Starship and the need of people with autism to travel into space.

Discussion

Summary of Findings

We found that after Musk's disclosure, the number of popular tweets about Asperger increased by almost 6-fold. This increase

XSL•FO
RenderX

in tweeting has the potential to promote awareness of the Asperger condition to more social media users. However, these "after" tweets seemed to receive less engagement than the ones posted before the disclosure.

Impact of a Celebrity Disclosure

As this was an observational study, the association between Musk´s disclosure and the increase in the number of tweets about Asperger receiving less engagement warrants further research to assess other possible factors affecting these results. It is nevertheless noteworthy that after Musk's disclosure, discussions about Asperger on Twitter increased. On May 9, 2021, Asperger became the 19th top trending topic on Twitter [33]. The interest in Asperger syndrome was also reflected in Google searches, when on May 11, 2021, Asperger ranked as the top 28th trending search [34]. This is remarkable since neither of the terms "Aspergers" or "Autism" ranked in the top 50 trending search topics during World Autism Awareness Day (April 2, 2021).

With the increase of popular posts about Asperger, both informative and misinformative tweets also appeared. Although nonsignificant, the few popular tweets that did contain misinformation tended to receive less engagement than posts that provided neutral or informative content. The lower engagement with tweets containing misinformation could suggest a kind of collective intelligence among Twitter users [35] that could contribute to increasing social knowledge by reducing the spread of misinformation as seen with the autism tweets.

Considering the impact that celebrity culture has on directing the public's attention to matters of health, it is worthwhile to investigate the types of information displayed on social media [36]. Within the field of autism research, very few studies assessing the informative and misinformative content posted on social media platforms have been conducted. This is slightly surprising since one of the more infamous studies, that resulted in a vast amount of misinformation on early childhood vaccination, was published (and later retracted) within this field. Misinformation about vaccines causing autism remains a problem today and is even propagated by some celebrities. Investigating the influence that celebrities, such as Musk, have on the public is underscored in our findings with the increase in attention that Asperger and autism received after Musk's disclosure.

Promoting the Awareness of Asperger and Autism Through Social Media

Promoting the awareness and acceptance of Asperger and autism by posting informative and factual content on social media platforms such as Twitter could assist in increasing social media users' knowledge about and understanding of the condition. Research on the use of social media for health promotion has shown its positive effects related to different conditions [15-20]. However, in our sample, we found that the tweets on Asperger that stimulated higher engagement tended to be neutral. These neutral posts did not misinform, but they did not provide any type of information that could increase one's knowledge or awareness about autism, either. Social media is ubiquitous and has become for most people a standard part of their everyday routine and habits. The ubiquitous use of social media platforms could provide an opportunity to expand the reach of trustworthy information about Asperger and autism in order to increase users' knowledge about and awareness of the conditions and reduce the associated stigmas [5,14,21].

Limitations

Our study had several limitations. We focused on a short period of time. We collected popular tweets about Asperger that were posted only in English. Furthermore, we only analyzed tweets posted the week after Musk's announcement, which may not be enough to form a conclusion on the effect of the disclosure. Also, Twitter users may not be representative of a random sample of the population, as the platform's users tend to range in age from 25 to 34 years [37]. Our findings may not apply to other social media platforms, to posts on autism that did not include our search terms, to posts in other languages, or to posts with less than 30 likes. We did not extract any identifiable information regarding the users that posted the popular tweets. Therefore, we cannot know if these tweets were posted by individuals or institutions. If posted by an institution, this could introduce a bias in terms of potential engagement impact (as institutions or organizations usually have more followers than individual accounts). Finally, due to the presence and sophistication of bot accounts, we cannot know if our sample size included tweets coming from any bot accounts [38].

Conclusions

The use of social media platforms by health authorities, autism associations, and other stakeholders has the potential to increase awareness of and knowledge about both Asperger and autism. When prominent figures disclose their conditions, such as autism, posts about their condition tend to increase, which provides an opportunity for trustworthy information about the condition to reach more social media users. Future research on autism and celebrity engagement with media should deploy in-depth data collection and longitudinal designs to detect possible changes in sentiment, as well as different methodological approaches (eg, qualitative, quantitative, and mixed designs) to elucidate underlying mechanisms at play related to the spread of both information and misinformation.

Conflicts of Interest

None declared.

References

1. World Health Organization. International statistical classification of diseases 11th revision: the global standard for diagnostic health information. World Health Organization. Geneva: WHO; 2018. URL: https://icd.who.int/en [accessed 2022-02-22]

2. Huynh S, McCrimmon A, Strong T. The change in classification of Asperger syndrome: an exploration of its effects on self-identity. Qual Rep 2020 Feb 14:379-397. [doi: 10.46743/2160-3715/2020.4122]

3. Kapp SK, Ne'eman A. Lobbying autism's diagnostic revision in the DSM-5. In: Kapp SK, editor. Autistic Community and the Neurodiversity Movement. London, UK: Palgrave Macmillan; 2020:167-194.

4. Kapp SK, editor. Autistic Community and the Neurodiversity Movement: Stories From the Frontline. London, UK: Palgrave Macmillan; 2020.

5. Skafle I, Gabarron E, Dechsling A, Nordahl-Hansen A. Online attitudes and information-seeking behavior on autism, Asperger syndrome, and Greta Thunberg. Int J Environ Res Public Health 2021 May 07;18(9):4981 [FREE Full text] [doi: 10.3390/ijerph18094981] [Medline: 34067114]

6. Dean M, Nordahl-Hansen A. A review of research studying film and television representations of ASD. Rev J Autism Dev Disord 2021 Jun 21:1-10. [doi: 10.1007/s40489-021-00273-8]

7. Hacking I. How we have been learning to talk about autism: a role for stories. Metaphilosophy 2009;20(3/4):499-516. [doi: 10.1111/j.1467-9973.2009.01607.x]

8. Nordahl-Hansen A, Oien R. Movie and TV depictions of autism spectrum disorder. In: Volkmar FR, editor. Encyclopedia of Autism Spectrum Disorders. New York, NY: Springer; 2017.

9. Nordahl-Hansen A, Tøndevold M, Fletcher-Watson S. Mental health on screen: a DSM-5 dissection of portrayals of autism spectrum disorders in film and TV. Psychiatry Res 2018 Apr;262:351-353. [doi: 10.1016/j.psychres.2017.08.050]

10. Skafle I, Nordahl-Hansen A, Øien RA. Short report: social perception of high school students with ASD in Norway. J Autism Dev Disord 2019 Nov 06;50(2):670-675. [doi: 10.1007/s10803-019-04281-w]

11. Howlin P, Goode S, Hutton J, Rutter M. Savant skills in autism: psychometric approaches and parental reports. Philos Trans R Soc Lond B Biol Sci 2009 May 27;364(1522):1359-1367 [FREE Full text] [doi: 10.1098/rstb.2008.0328] [Medline: 19528018]

12. Wood C, Freeth M. Students' stereotypes of autism. JEI 2016 Oct 18;2(2):131. [doi: 10.5296/jei.v2i2.9975]

13. Yu L, Stronach S, Harrison AJ. Public knowledge and stigma of autism spectrum disorder: comparing China with the United States. Autism 2020 Apr 27;24(6):1531-1545. [doi: 10.1177/1362361319900839]

14. Morgan AJ, Ross A, Reavley NJ. Systematic review and meta-analysis of Mental Health First Aid training: effects on knowledge, stigma, and helping behaviour. PLoS ONE 2018 May 31;13(5):e0197102. [doi: 10.1371/journal.pone.0197102]

15. An R, Ji M, Zhang S. Effectiveness of social media-based interventions on weight-related behaviors and body weight status: review and meta-analysis. Am J Health Behav 2017 Nov 01;41(6):670-682. [doi: 10.5993/AJHB.41.6.1] [Medline: 29025495]

16. Dimanlig-Cruz S, Han A, Lancione S, Dewidar O, Podinic I, Haqani B, et al. Physical distancing messages targeting youth on the social media accounts of Canadian public health entities and the use of behavioral change techniques. BMC Public Health 2021 Sep 07;21(1):1634-1634. [doi: 10.1186/s12889-021-11659-y]

17. Gabarron E, Wynn R. Use of social media for sexual health promotion: a scoping review. Glob Health Action 2016;9:32193 [FREE Full text] [doi: 10.3402/gha.v9.32193] [Medline: 27649758]

18. Gabarron E, Årsand E, Wynn R. Social media use in interventions for diabetes: rapid evidence-based review. J Med Internet Res 2018 Aug 10;20(8):e10303 [FREE Full text] [doi: 10.2196/10303] [Medline: 30097421]

19. Hudnut-Beumler J, Po'e E, Barkin S. The use of social media for health promotion in Hispanic populations: a scoping systematic review. JMIR Public Health Surveill 2016 Jul 11;2(2):e32 [FREE Full text] [doi: 10.2196/publichealth.5579] [Medline: 27400979]

20. Simeon R, Dewidar O, Trawin J, Duench S, Manson H, Pardo Pardo J, et al. Behavior change techniques included in reports of social media interventions for promoting health behaviors in adults: content analysis within a systematic review. J Med Internet Res 2020 Jun 11;22(6):e16002 [FREE Full text] [doi: 10.2196/16002] [Medline: 32525482]

21. Calhoun AJ, Gold JA. "I feel like I know them": the positive effect of celebrity self-disclosure of mental illness. Acad Psychiatry 2020 Feb 25;44(2):237-241. [doi: 10.1007/s40596-020-01200-5]

22. Kresovich A, Noar SM. The power of celebrity health events: meta-analysis of the relationship between audience involvement and behavioral intentions. J Health Commun 2020 Sep 29;25(6):501-513. [doi: 10.1080/10810730.2020.1818148]

23. Indratmo Z, Buro K. Comparisons between text-only and multimedia tweets on user engagement. 2020 Presented at: 2020 IEEE International Conference on Systems, Man, and Cybernetics; October 11-14, 2020; Toronto, Canada p. 3825-3831. [doi: 10.1109/smc42975.2020.9283438]

24. Kaur W, Balakrishnan V, Rana O, Sinniah A. Liking, sharing, commenting and reacting on Facebook: user behaviors' impact on sentiment intensity. Telemat Inform 2019 Jun;39:25-36. [doi: 10.1016/j.tele.2018.12.005]

25. Gabarron E, Dechsling A, Skafle I, Nordahl-Hansen A. Repository: Aspergers gaining momentum on social media. Zenodo. 2021. URL: https://doi.org/10.5281/zenodo.5656916 [accessed 2022-02-21]

26. Nielsen F. A New ANEW: evaluation of a word list for sentiment analysis in microblogs. 2011 Presented at: ESWC2011 Workshop on 'Making Sense of Microposts'; May 30, 2011; Heraklion, Greece. [doi: 10.1117/2.5200207.0007]

XSL•FO
RenderX

27.   Marcec R, Likic R. Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines. Postgrad Med J 2021 Aug 09:postgradmedj-2021-140685. [doi: 10.1136/postgradmedj-2021-140685]

28.   Borakati A. Evaluation of an international medical E-learning course with natural language processing and machine learning. BMC Med Educ 2021 Mar 25;21(1):181-181. [doi: 10.1186/s12909-021-02609-8]

29.   Denecke K, Gabarron E. How artificial intelligence for healthcare look like in the future? Stud Health Technol Inform 2021 May 27;281:860-864. [doi: 10.3233/SHTI210301] [Medline: 34042796]

30.   Compagner C, Lester C, Dorsch M. Sentiment analysis of online reviews for selective serotonin reuptake inhibitors and serotonin-norepinephrine reuptake inhibitors. Pharmacy (Basel) 2021 Jan 23;9(1):27 [FREE Full text] [doi: 10.3390/pharmacy9010027] [Medline: 33498697]

31.   Yoon S, Parsons F, Sundquist K, Julian J, Schwartz JE, Burg MM, et al. Comparison of different algorithms for sentiment analysis: psychological stress notes. Stud Health Technol Inform 2017;245:1292 [FREE Full text] [Medline: 29295377]

32.   Durand WM, Peters JL, Eltorai AEM, Kalagara S, Osband AJ, Daniels AH. Medical crowdfunding for organ transplantation. Clin Transplant 2018 May 15;32(6):e13267. [doi: 10.1111/ctr.13267]

33.   Twitter trends on 9th May, 2021. Trend Calendar. 2021. URL: https://us.trend-calendar.com/trend/2021-05-09.html [accessed 2022-02-21]

34.   Google trends on 11th May, 2021. Trend Calendar. URL: https://us.trend-calendar.com/trend/2021-05-11.html [accessed 2022-02-21]

35.   Becker J, Brackbill D, Centola D. Network dynamics of social influence in the wisdom of crowds. Proc Natl Acad Sci USA 2017 Jun 12;114(26):201615978. [doi: 10.1073/pnas.1615978114]

36.   Hoffman SJ, Tan C. Following celebrities' medical advice: meta-narrative analysis. BMJ 2013 Dec 17;347(dec17 14):f7151-f7151. [doi: 10.1136/bmj.f7151]

37.   Statista. Distribution of Twitter users worldwide as of April 2021, by age group. 2021. URL: https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/ [accessed 2022-02-21]

38.   Antonakaki D, Fragopoulou P, Ioannidis S. A survey of Twitter research: data model, graph structure, sentiment analysis and attacks. Expert Syst Appl 2021 Feb;164:114006. [doi: 10.1016/j.eswa.2020.114006]

Abbreviations

DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition

<u>Original Paper</u>

A Tailored App for the Self-management of Musculoskeletal Conditions: Evidencing a Logic Model of Behavior Change

Alice Berry[1], BSc (Hons), MSc, PhD; Carey McClellan[2], BSc (Hons), PhD; Ben Wanless[3], BSc (Hons), MSc; Nicola Walsh[1], BSc, MSc, PhD

[1]University of the West of England, Bristol, United Kingdom

[2]getUBetter, Bristol, United Kingdom

[3]Musculoskeletal Department, St George's University Hospitals National Health Service Foundation Trust, London, United Kingdom

Corresponding Author:
Alice Berry, BSc (Hons), MSc, PhD
University of the West of England
Glenside Campus
Blackberry Hill
Bristol, BS16 1DD
United Kingdom
Phone: 44 117 32 87802
Email: alice.berry@uwe.ac.uk

Abstract

Background: Musculoskeletal conditions such as joint pain are a growing problem, affecting 18.8 million people in the United Kingdom. Digital health interventions (DHIs) are a potentially effective way of delivering information and supporting self-management. It is vital that the development of such interventions is transparent and can illustrate how individual components work, how they link back to the theoretical constructs they are attempting to change, and how this might influence outcomes. getUBetter is a DHI developed to address the lack of personalized, supported self-management tools available to patients with musculoskeletal conditions by providing knowledge, skills, and confidence to navigate through a self-management journey.

Objective: The aim of this study was to map a logic model of behavior change for getUBetter to illustrate how the content and functionality of the DHI are aligned with recognized behavioral theory, effective behavior change techniques, and clinical guidelines.

Methods: A range of behavior change models and frameworks were used, including the behavior change wheel and persuasive systems design framework, to map the logic model of behavior change underpinning getUBetter. The three main stages included understanding the behavior the intervention is attempting to change, identifying which elements of the intervention might bring about the desired change in behavior, and describing intervention content and how this can be optimally implemented.

Results: The content was mapped to 25 behavior change techniques, including information about health consequences, instruction on how to perform a behavior, reducing negative emotions, and verbal persuasion about capability. Mapping to the persuasive system design framework illustrated the use of a number of persuasive design principles, including tailoring, personalization, simulation, and reminders.

Conclusions: This process enabled the proposed mechanisms of action and theoretical foundations of getUBetter to be comprehensively described, highlighting the key techniques used to support patients to self-manage their condition. These findings provide guidance for the ongoing evaluation of the effectiveness (including quality of engagement) of the intervention and highlight areas that might be strengthened in future iterations.

(JMIR Form Res 2022;6(3):e32669) doi:10.2196/32669

KEYWORDS

musculoskeletal; supported self-management; behavior change; digital health intervention; behavior change wheel

XSL•FO
RenderX

Introduction

Background

Musculoskeletal conditions such as joint pain are a growing problem, affecting 18.8 million people in the United Kingdom, which equates to 22% of the total burden of ill health across the population [1]. The importance of supporting people with musculoskeletal conditions in self-managing their own condition is increasingly acknowledged as a valuable tool for improving symptoms and facilitating healthy behaviors [2].

Self-management includes all the actions taken by people to recognize, treat, and manage their own health care independently of, or in partnership with, the health care system [3]. Supported self-management requires a whole systems approach that should be fully integrated into clinical care pathways, providing information and encouragement to help people gain more control over symptoms by understanding their condition through monitoring and taking appropriate action [4].

Adopting digital methods to support the self-management of a number of health conditions has become increasingly popular, gaining further traction during the COVID-19 pandemic with a rise in the number of digital health consultations [5]. Recent literature has highlighted the benefit of using digital technology in general practice, suggesting that the National Health Service (NHS) continues to invest in digital-first primary care [6].

Despite the rise in popularity, it is recognized that most digital health interventions (DHIs) are not well-described [7]. It is vital that the development of DHIs is transparent and can illustrate how individual components work, how they link back to the theoretical constructs they are attempting to change, and how this might influence outcomes [8]. During DHI development, it is important to clearly document techniques used by mapping content against recognized behavior change taxonomies [7]. This approach can allow for better implementation into practice, producing more effective and longer term engagement with healthy behaviors [9]. The National Institute for Health and Care Excellence (NICE) Evidence Standards Framework for digital health technology (DHT) also highlights the importance of using recognized effective behavior change techniques (BCTs), recommending that evidence is available to illustrate how BCTs within an intervention are consistent with recognized behavior change theory and recommended practice (aligned with guidance from NICE or relevant professional organizations) and appropriate for the target population [10].

As part of the Small Business Research Initiative Health Care and an NHS England program to bring new technologies to the NHS, getUBetter was contracted to develop, deploy, evaluate, and optimize its DHT to scale into the NHS. The contract was titled *Integrated Self-management and Prevention for Multiple Musculoskeletal Conditions and Pathways*. getUBetter worked with teams from the Wandsworth NHS Clinical Commissioning Group, St Georges University Hospitals NHS Foundation Trust, the Digital Evaluation Unit at the Health Innovation Network in South London, and the University of the West of England in Bristol to evaluate the DHI in a real-world setting. Part of this evaluation focused on evidencing the underpinning logic model of behavior change in the DHI; this element of the project is reported in this paper.

getUBetter [11] was developed to address the recognized problem within the health system of a lack of personalized, supported self-management tools available to patients with musculoskeletal conditions that provide the knowledge required to trust in, and successfully navigate through, a self-management journey.

The aim of the getUBetter self-management app is to empower individuals to self-manage from the start of an acute phase of an injury or episode to recovery and long-term condition management. It intends to do this by providing knowledge, developing skills, identifying and developing support strategies, and using persuasive language to develop positive beliefs about one's self-management capability. Patients are prescribed getUBetter by a clinician (or practice staff) face to face, virtually, or via an SMS text message as part of routine care. Patients can also self-refer from a health care provider's website. Musculoskeletal conditions include back, back and leg, neck, shoulder, knee, ankle, and lower limb soft injury, and patients can be linked to ≥1 local injury or condition pathway simultaneously, such as lower back pain and new ankle injury.

getUBetter was designed for both new and recurrent conditions and is specifically tailored to local care pathways, supporting triage, day-to-day recovery, referral when needed, connection to local rehabilitation programs, and prevention and well-being when a patient is back to their normal.

The getUBetter platform *houses* standardized national care pathways for the range of conditions based on evidence and national clinical treatment guidelines. Before the getUBetter platform and app are integrated into a health care system (eg, into general practitioner practices), urgent care centers, or physiotherapy departments, getUBetter works with the local clinical teams to configure each element of the local care pathway. There are three pillars of configurability: the day-by-day condition (lived experience of the patient), the end-to-end clinical condition pathways, and the local health system (health care and social care and public health). Patients are then connected to their specifically tailored local care pathways via either prescriptions or self-referrals.

getUBetter adopts a whole pathway approach that includes a symptoms checker, use of stratification tools, and automated referrals. Targeted and personalized content aims to help patients understand their stage of recovery, providing knowledge to strengthen confidence; reduce negative behaviors; and include a recovery plan around medication, effective treatment options, and the ability to self-monitor recovery. Evidence-based advice is provided on a range of topics such as what to expect at different stages of recovery; information about pain relief; guidance on referral options; information about driving, work, and other everyday activities; a symptom checker; information about red flags; and a variety of exercises demonstrated by physiotherapists are available to guide patients through their recovery journey (for more examples, see Multimedia Appendix 1 [12]).

The co-design process included patient and public involvement groups from the local clinical commissioning group, general practitioners, physiotherapists of all grades, consultant or advanced practice physiotherapists for most conditions, carers, charities such as Age UK, and patients. Patients from the patient and public involvement group were a broad group with differing technical capabilities, and the focus groups comprised patients of all ages. Importantly, ongoing user experience was captured on a regular basis during development.

getUBetter is suitable for all patients aged ≥18 years. The platform and app have been co-designed with all stakeholder groups to minimize digital exclusion; for example, easy-to-use buttons, text size to 200%, and an easy-to-access web version. getUBetter works closely with organizations, patients, and all stakeholders to minimize digital exclusion and support *care as usual* and actively seeks to support patients who may be having difficulties accessing the solution.

Aim of Study

The aim of this study was to develop a logic model of behavior change for getUBetter, illustrating how the content and functionality of the app are aligned with recognized behavioral theory, effective techniques, and recommended clinical practice. This paper describes the development of a behavior change model for getUBetter using widely used models and frameworks described in more detail in the following sections.

Methods

Overview

The behavior change wheel (BCW) [13], Capability Opportunity Motivation-Behavior (COM-B) model [14], and BCT Taxonomy (BCTT) Version 1 (BCTTv1) [7] were used to map the BCTs present and the behavior change model for getUBetter. The persuasive systems design (PSD) framework [15] was used to describe the functionality and delivery of content.

To replicate and implement behavior change interventions in practice, we need an agreed-upon language to report their content or *active ingredients*. There are a range of tools that can be used to do this, and one that is widely used is the BCW, which provides a systematic way of characterizing interventions. It is often used in conjunction with the COM-B framework [14], a theoretical framework that incorporates key components (capability, opportunity, and motivation) considered to affect behavior. The COM-B model is a starting point for understanding behavior in the context in which it occurs. The central tenet of the model is that for any behavior to occur (1) there must be the *capability* to do it (ie, physical strength, knowledge, skills, and stamina), (2) there must be the *opportunity* for the behaviors (ie, physically accessible, affordable, socially acceptable, and sufficient time), and (3) there must be sufficiently strong *motivation* [12]. BCTTv1 was developed to specify content in terms of BCTs, the smallest components of behavior change interventions that can bring about change [7]. BCTTv1 is a hierarchically structured

taxonomy of 93 distinct BCTs, including labels, definitions, and examples.

Although the BCTT enables the illustration of BCTs adopted within an intervention, the additional inclusion of the PSD framework was important to demonstrate how BCTs were translated into practice and ensure that important aspects of functionality were adequately described. The PSD framework was developed to guide the process of designing and evaluating persuasive systems, which highlights 28 design principles for persuasive system content and functionality, categorized into four areas: primary task, dialog, system credibility, and social support categories [15].

The following sections describe the 3 stages that were conducted to illustrate the behavior change model for getUBetter. It is important to highlight that this process was nonlinear, and *mapping* procedures moved between stages iteratively to link BCTs with theory, intended outcomes, functions, and categories of the BCW, as well as persuasive system design elements.

Stage 1: Understanding the Target Behavior

Several team meetings comprised clinicians, developers, user experience and user interface designers, project managers, and product owners from getUBetter alongside other clinicians from the stakeholder groups and researchers with experience in developing self-management and behavior change interventions from the University of the West of England. The clinicians included an advanced physiotherapy practitioner in primary care and emergency care and a consultant musculoskeletal physiotherapist. The meetings were held to understand the health problems that getUBetter aims to address and the behavior(s) that it attempts to change and describe what needs to change to achieve the intended outcome(s). This process enabled the aims and objectives of getUBetter to be comprehensively described and illustrated in a logic model of the problem.

Next, the COM-B model and the Theory and Technique tool [16] were used to link the BCTs present in getUBetter to relevant theoretical domains. The Theory and Technique tool links the theoretical domains framework [17] to the COM-B model to provide a more granular understanding of how constructs are associated with capability, opportunity, and motivation (eg, the theoretical domain *beliefs about capabilities* is linked to *reflective motivation* within the COM-B model). This step highlighted to the team the potential areas where getUBetter might be strengthened by adding in BCTs commonly reported in similar published literature but which were not currently present in getUBetter.

Stage 2: Mapping Intervention Options

This stage involved linking information gathered previously in stage 1 (the *behavioral diagnosis*) to the intervention functions (Figure 1) and policy categories (Figure 1) through which getUBetter is currently being implemented. Team meetings were held to agree on the links between the BCTs present in getUBetter and relevant implementation functions and policy categories. Once again, this was an opportunity to explore the areas that could be strengthened in future iterations.

Figure 1. Behavior change wheel [13].
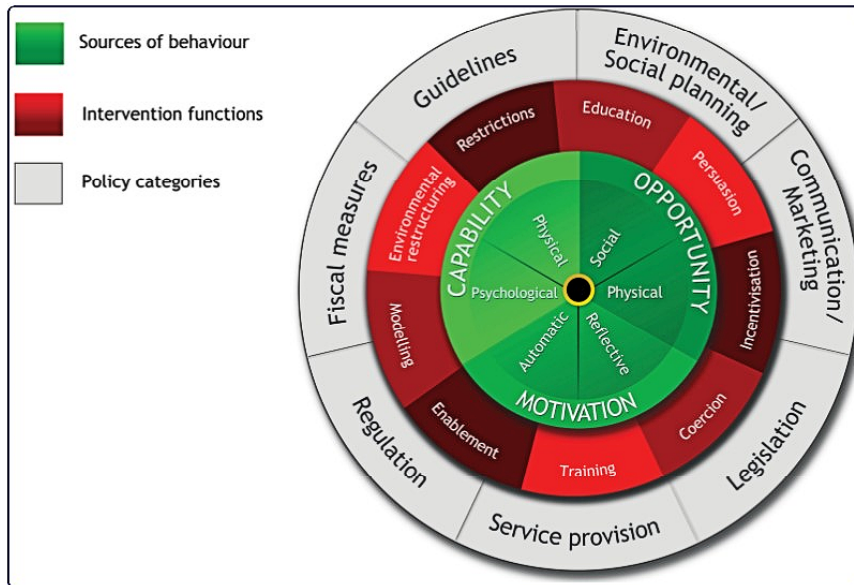
Stage 3: Mapping Content and Implementation Options

This final stage involved using the BCTT [7] to identify and map BCTs present within getUBetter and the PSD model to illustrate how the BCTs were being delivered (Figure 2).

Figure 2. Persuasive systems design model.

The results from the abovementioned stages described were then consolidated into a logic model of behavior change for getUBetter.
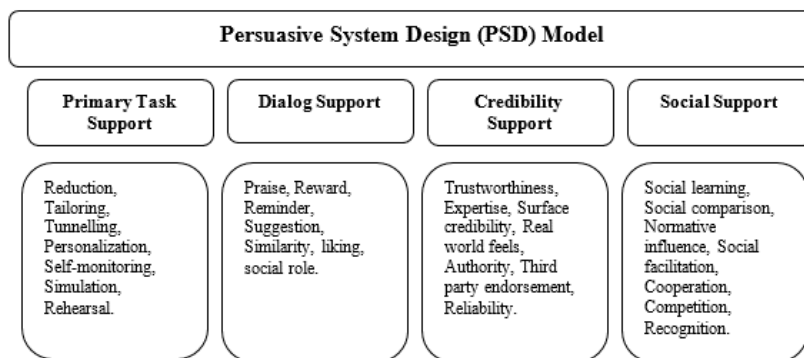
Results

Overview

An initial mapping exercise was completed during November 2019 and April 2020 by the lead author (AB), who discussed and agreed on the mapping with team members. The mapping process was nonlinear and iterative, moving between the stages described earlier and guided by team discussions. This initial mapping exercise enabled the team to identify particular areas of the app that could be strengthened by adding specific BCTs commonly used in other interventions [12] but which were not currently present in getUBetter. Following ongoing iterative design and development of getUBetter throughout 2020, a second mapping exercise took place between January 2021 and March 2021 to highlight additional techniques and functionalities.

The following sections describe each stage of *mapping round 1* in more detail and the additional techniques and changes in functionality highlighted during *mappinground 2*.

Stage 1: Understanding the Target Behavior

Overview

The aim of getUBetter is to empower individuals to self-manage their musculoskeletal condition by providing support and advice throughout their recovery journey and prevent future episodes. Objectives were clustered into three subgroups: knowledge and skills, supportive environment, and motivation and self-efficacy. The aims and objectives are described in more detail in the following sections.

Behavioral Aims and Objectives of getUBetter (version 2.0)

Knowledge and Skills

These objectives aim to provide the patient with sufficient knowledge about their musculoskeletal condition, enabling them to develop sufficient skills to conduct positive self-management behaviors.

The objectives were as follows:

1. Educate about normal physiological responses (ie, symptoms and time frames)
2. Train the patient in personally relevant skills (such as reducing sedentary behavior, introducing graded tasks, and relevant exercises) to enable them to manage symptoms and implement positive self-management behaviors

Supportive Environment

These objectives aim to train the patient to identify and develop supportive aspects of physical and social environments.

The objectives were as follows:

1. Guide understanding of the interpersonal influences on self-managing their musculoskeletal condition and how the environment can be restructured to provide optimal support
2. Promote reflection on environmental and emotional aspects affecting self-management (time, resource use, and place and well-being)
3. Highlight opportunities for restructuring the physical and social environments to best support self-management behaviors
4. Enable patient to navigate the health system and understand treatment pathways and referral options

Motivation and Self-efficacy

These objectives aim to enable the patient to form positive beliefs and feelings toward self-managing their musculoskeletal condition and support the formation of a personally relevant action plan for self-management behaviors

The objectives were as follows:

1. Persuade the patient about their capability and enable them to perform self-management behaviors such as action planning or healthy lifestyles
2. Educate (to strengthen beliefs) about the positive consequences (outcomes) of the directed self-management behaviors
3. Train the patient to form clear action plans for conducting self-management behaviors and self-monitor behaviors and outcomes
4. Provide prompts and cues to encourage self-management behaviors.

Specific self-management behaviors were also identified (intended outcomes; Figure 3). Figure 3 shows the logic model of the problem(s) that getUBetter aims to address, how it aims to engage patients with positive self-management behaviors, and the intended outcomes.

The next step was to identify how BCTs present in getUBetter were linked back to the theoretical constructs they were attempting to change (using the BCW guidance and the Technique and Theory Tool see *Methods* section). Textbox 1 illustrates these links.

This stage highlighted a number of BCTs that could be beneficial to include in future iterations of getUBetter. More specifically, BCW guidance highlighted that BCTs related to social comparison (5.2) and goal setting and monitoring (ie, 1.1, 1.5, and 2.3) were reported to have been used in previous behavior change interventions [12].

Figure 3. Logic model of problem.
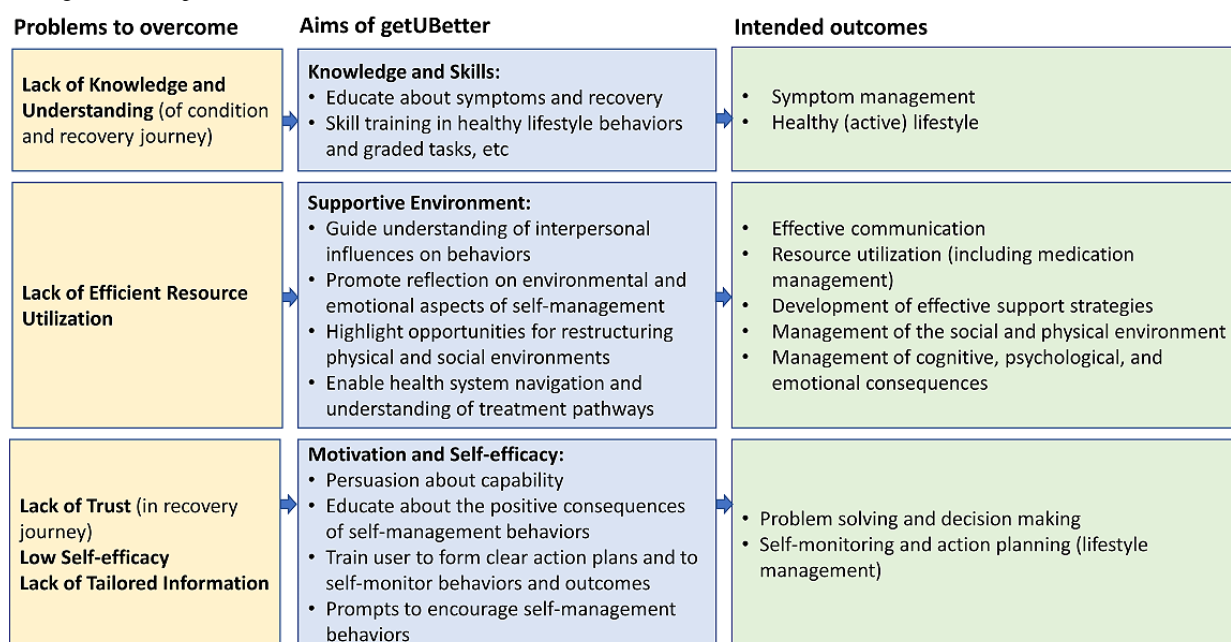
Textbox 1. Capability, Opportunity, and Motivation–Behavior analysis (coded to Behavior Change Technique Taxonomy version 1).

Capability

- Knowledge and skills

  - 4.1 Instruction on how to perform a behavior

  - 5.1 Information about health consequences

  - 8.1 Behavioral practice and rehearsal

  - 8.7 Graded tasks

- Memory, attention, and decision processes

  - 7.1 Prompts and cues

- Behavioral regulation

  - 1.2 Problem solving

  - 8.2 Behavior substitution

  - 11.2 Reduce negative emotions

Opportunity

- Environmental context and resources

  - 7.1 Prompts and cues

  - 12.1 Restructuring the physical environment

  - 12.2 Restructuring the social environment

- Social influences and identity

  - 3.1 Social support (unspecified)

  - 6.2 Social comparison

  - 9.1 Credible source

Motivation

- Beliefs about capabilities

  - 1.2 Problem solving

  - 4.1 Instruction on how to perform a behavior

  - 6.1 Demonstration of the behavior

  - 8.1 Behavioral practice and rehearsal

  - 8.7 Graded tasks

  - 15.1 Verbal persuasion about capability

- Beliefs about consequences

  - 5.1 Information about health consequences

  - 5.6 Information about emotional consequences

- Intentions

  - 5.1 Information about health consequences

- Reinforcement

  - 7.1 Prompts and cues

- Emotion

  - 11.2 Reduce negative emotions

XSL•FO

RenderX

Behavioral Support at Different Stages of Recovery

The mapping process described in this paper provided an opportunity for the team to highlight that a key aim of getUBetter is to empower individuals to self-manage their condition wherever they may be on their recovery journey. This issue was not governed by the mapping process; however, the team agreed that it was important to describe this during the mapping process to illustrate the nuance of delivering BCTs in both a personalized and tailored way. The team also consulted with clinicians at this point to gather opinions on the potential weighting of BCTs at different stages of recovery. Although there were some differing opinions, there was general agreement in line with the weighting shown in Figure 4.

Figure 4 illustrates how, within getUBetter, content is tailored to two main variables: (1) timeline of the recovery journey, moving through early-term, mid-term, and long-term stages of recovery, and (2) status—how the patient is progressing at any given time point (ie, getting better, staying the same, or getting worse). Content is also personalized to two main variables: (1) an individual's local care pathway, local services, and health care provider and (2) patients identifying their main problems and worries to receive personalized advice. Content automatically adjusts to the stage of the recovery journey and individual progress, and this targeted content helps the patient to understand their stage of recovery while providing personalized support. Figure 4 illustrates how content is both tailored and personalized.

The full range of BCTs is present across all stages of recovery, although weighted differently according to individual progress and stage of recovery. For example, in the earlier stages of recovery, the key focus is on reassurance and providing knowledge about what to expect, what is normal, and having the confidence to seek help if needed. Later on, the focus might shift to making an action plan and considering healthy lifestyle changes to support effective self-management (Figure 4 also illustrates how content is weighted at different stages of recovery). The main concerns and problems can also be identified to further personalize content to strengthen confidence, reduce negative behaviors, and provide ongoing reassurance and persuasion about conducting self-management behaviors.

Figure 4. Support at different stages of recovery.
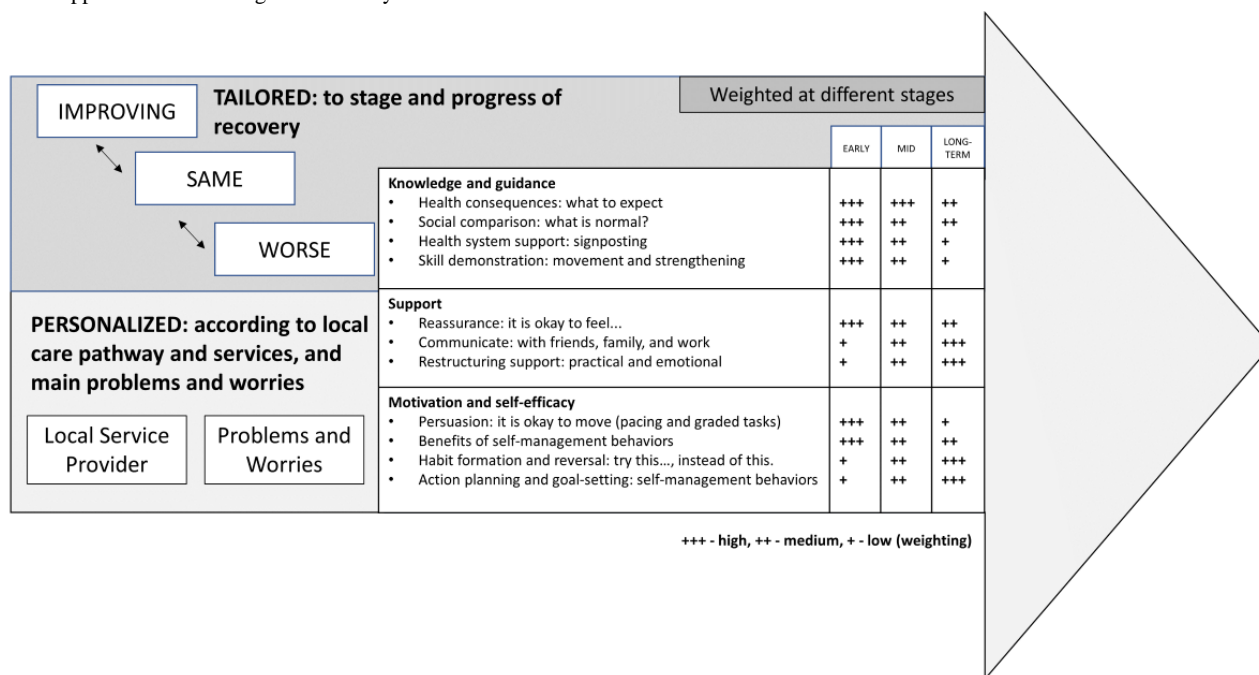
Stage 2: Mapping Intervention Options

BCTs present in getUBetter were then mapped to the intervention functions and policy categories of the BCW. This was another opportunity to evaluate how well the existing BCTs in getUBetter represented each intervention function.

Relevant intervention functions included education, persuasion, training, environmental restructuring, modeling, and enablement. Policy categories relevant to support the delivery of getUBetter included communication and marketing, guidelines, fiscal measures, regulation, service provision, and environmental and social planning. Multimedia Appendix 2 provides a detailed description of the links between the COM-B components and the intervention functions.

Stage 3: Mapping Content and Implementation Options

Mapping: Round 1

BCT Mapping

During the first round of mapping, getUBetter was mapped to 22 BCTs in the BCTTv1 (Textbox 2).

Textbox 2. Mapping of behavior change techniques in getUBetter.

Behavior change technique (coded to Behavior Change Technique Taxonomy version 1)

- 1.1 Goal setting (behavior; *added in round 2*)

- 1.2 Problem solving

- 1.5 Review behavior goal(s) (*added in round 2*)

- 2.3 Self-monitoring of behavior (*added in round 2*)

- 2.5 Monitoring of outcome(s) of behavior without feedback

- 3.1 Social support (unspecified)

- 3.3 Social support (emotional)

- 4.1 Instruction on how to perform a behavior

- 5.1 Information about health consequences

- 5.6 Information about emotional consequences

- 6.1 Demonstration of the behavior

- 6.2 Social comparison

- 7.1 Prompts and cues

- 8.1 Behavioral practice and rehearsal

- 8.2 Behavior substitution

- 8.3 Habit formation

- 8.4 Habit reversal

- 8.7 Graded tasks

- 9.1 Credible source

- 11.1 Pharmacological support

- 11.2 Reduce negative emotions

- 12.1 Restructuring the physical environment

- 12.2 Restructuring the social environment

- 12.6 Body changes

- 15.6 Verbal persuasion about capability

PSD Elements

Mapping to the PSD model illustrated the use of a number of persuasive design principles, including tailoring, personalization, simulation, and reminders, which were used to deliver the content of getUBetter.

Table 1 below shows how getUBetter incorporates key persuasive design principles from the PSD model into its functionality and content. In addition, it was recognized that some of these might be adapted and strengthened in future iterations.

Table 1. Persuasive system design elements in getUBetter.

Elements and descriptions	Existing content in getUBetter
Primary task support	
Reduction	
A system that reduces complex behavior into simple tasks helps patients perform the target behavior, and it may increase the benefit–cost ratio of a behavior.	• Simplification of national guidelines and musculoskeletal pathway • Simplifies the journey through local pathway and access to local services • Automates referrals and promotes self-choice • Writing style—translates complex messages into simple language • Simplifies tasks—related to time point
Tunneling	
Using the system to guide patients through a process or experience provides opportunities to persuade along the way.	• Tunneled by timeline—content provided depending on choice: improving, staying the same, or getting worse • Full autonomy for which order to visit app pages • Symptom checker—red flags directing the patient to a health professional (lock screen)
Tailoring	
Information provided by the system will be more persuasive if it is tailored to the potential needs, interests, personality, use context, or other factors relevant to a patient group.	• All content tailored to the day of recovery and recovery status (getting better, staying the same, or getting worse) • Feedback zone—tailored information around main worries and problems
Personalization	
A system that offers personalized content or services has a greater capability for persuasion.	• My main problems • My main worries (provides tailored information) • Personalized information to local services (local trust specific)
Self-monitoring	
A system that keeps track of one's own performance or status supports the patient in achieving goals.	• Diary, calendar, and goal-setting functions are available in the app, which are linked to prompts and reminders
Simulation	
Systems that provide simulations can persuade by enabling patients to observe the link between cause and effect immediately.	• Videos about simple exercises and activities of daily living
Rehearsal	
A system providing means with which to rehearse a behavior can enable people to change their attitudes or behavior in the real world.	• Videos and instructions on how to perform a behavior—exercising specific and general day-to-day activities
Dialogue support	
Praise	
The system should use praise via words, images, symbols, or sounds as a way of providing patient feedback information based on their behaviors.	• Persuasive messaging throughout (including information about health consequences, reducing negative emotions, and verbal persuasion about the capability to perform self-management behaviors—covered in BCTTv1[a] mapping)
Reminders	
If a system reminds patients of their target behavior, the patients will more likely achieve their goals. The system should remind patients of their target behavior during the use of the system.	• Automated reminders to log back into the app
Suggestion	
The system should suggest that patients conduct behaviors during the system use process.	• Covered in BCTTv1 mapping
Liking	
A system that is visually attractive for its patients is likely to be more persuasive.	• Ongoing design updates

XSL•FO
RenderX

Elements and descriptions	Existing content in getUBetter
System credibility support	
Trustworthiness	
The system should provide information that is truthful, fair, and unbiased.	• Follows national guidelines, NHS[b] guidelines, and NICE[c] digital health guidelines and is validated by CCG[d]—linked to the musculoskeletal pathway • Local GP[e] logo
Expertise	
The system should provide information showing knowledge, experience, and competence.	• Developed by experts
Surface credibility	
The system should have a competent look and feel.	• P-mat—tested with patients—high level of usability and acceptability • No advertisements
Real-world feel	
The system should provide information about the organization and actual people behind its content and services.	• NHS badge, local services, local CCG—credible source • In GP practices
Authority	
The system should refer to people in the role of authority.	• Links back to GP and other NHS services
Third-party endorsements	
The system should provide endorsements from respected sources.	• Links back to GP and other NHS services
Verifiability	
The system should provide means of verifying the accuracy of the site content via outside sources.	• Links back to GP and other NHS services
Social support	
Social learning	
The system should provide means of observing other patients who are performing their target behaviors and seeing the outcomes of their behavior.	• Some target behaviors were displayed in videos but not all • No illustrations of outcomes of behavior (eg, positive stories of improvement)

[a]BCTTv1: Behavior Change Technique Taxonomy Version 1.

[b]NHS: National Health Service.

[c]NICE: National Institute for Health and Care Excellence.

[d]CCG: clinical commissioning group.

[e]GP: general practitioner.

Mapping: Round 2

The first round of mapping highlighted the potential value of adding specific BCTs to strengthen it in future iterations. Following a number of changes during ongoing development, a second round of mapping was conducted, resulting in the identification of 3 additional BCTs (Textbox 2). This was translated to a number of areas within the app that were strengthened; for example, the addition of a goal-setting section linked to a calendar function where patients can log progress and self-monitor their recovery journey.

A number of PSD elements were also strengthened, including more direct links between sections, production of additional guidance and instruction videos, development of video strings to guide patients through videos, and addition of a traffic light system to guide the patient to the most appropriate exercises.

Logic Model of Behavior Change for getUBetter

Figure 5 illustrates the full behavior change model for getUBetter. It provides details of the target COM-B elements that getUBetter intends to change (target behaviors) and describes the intervention functions and BCT groups that are used to make this change (description of intervention), the PSD elements present that deliver the BCTs (mode of delivery) via the app, and how the intervention as a whole affects the intended outcomes (outputs).

Figure 5. Logic model of behavior.
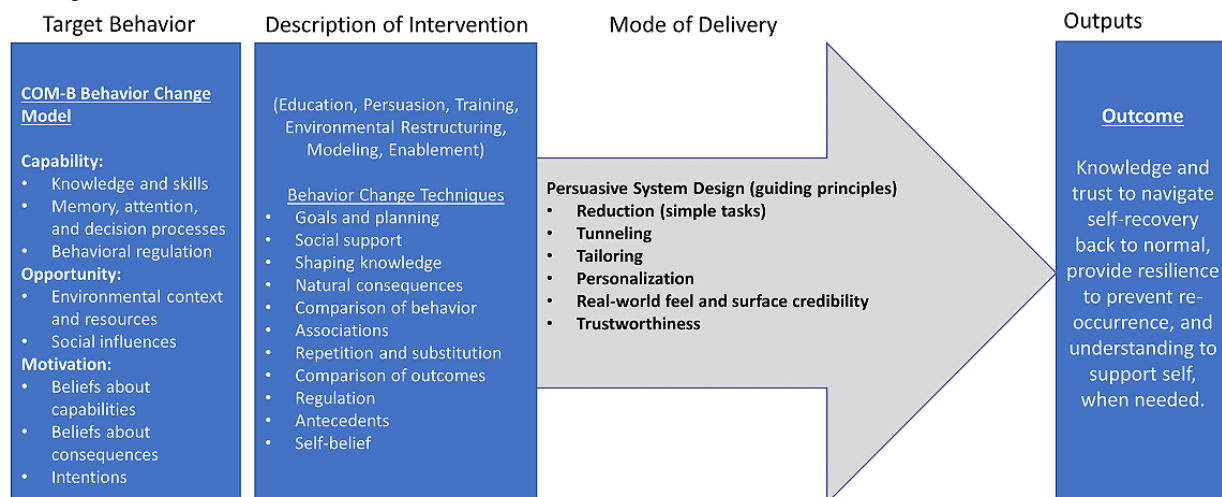
Discussion

Principal Findings

The aim of this project was to evidence the behavior change model underpinning getUBetter. A range of models and frameworks were adopted to demonstrate how the content and functionality of the intervention aligns with the recognized behavioral theory, effective BCTs, and recommended clinical practice.

This study offers a method for illustrating a behavior change model for a complex DHI and is in line with Medical Research Council complex intervention development guidelines [18], NICE evidence standards framework for DHT [10], and literature that has adopted similar methods to illustrate the development of behavior change interventions in musculoskeletal health [19]. It recognizes the value of using overlapping frameworks to capture recognized BCTs (using the BCTTv1) while highlighting the novel functionality of the intervention (using the PSD model), which importantly enables transparency and replicability.

The behavior change model can be seen as a blueprint for behavior change that can be used to guide future iterations of the intervention and the development of content for other musculoskeletal conditions, which are currently in development. The process also enabled the team to identify potential areas of the intervention that could be strengthened and highlighted the potential value of adding BCTs not present in the initial mapping round.

The mapping exercise opened up an opportunity for team discussion about the relative importance of different BCTs at different time points of recovery. getUBetter was developed so that content is tailored depending on where each individual is within their own recovery journey and whether symptoms are improving, staying the same, or getting worse. Future research should explore with patients the relative importance of different behavioral support at different stages of recovery to help identify if and how behavioral support can be more tailored and personalized in future iterations of the DHI. This is a novel area

that is yet to be explored in the literature and requires further investigation.

Limitations

Mapping and coding of BCTs are often open to interpretation, and we recognize the potential bias from the project team. However, a number of team meetings (which included clinicians with extensive experience in treating musculoskeletal conditions and academics with experience in developing behavior change interventions) were held to discuss the content and reach consensus when linking content to BCTs.

Coproduction and engagement with patients (a planned next step) would add value and help to learn more about potential areas that could be better represented and additional BCTs relevant and acceptable to the musculoskeletal population.

Next Steps

Throughout this process, the importance of exploring acceptability (of intervention content) and usability (of the digital intervention) was recognized as a key factor in understanding how the intervention will be adopted and used by patients, and work in this area is being planned in future studies. The behavior change model described in this paper will act as a useful foundation for exploring these topics further with patients, and we recognize the important need to understand the characteristics and strategies that are most effective in promoting both the engagement with digital technologies [20] and the healthy behaviors that they intend to promote. Future work will also explore the effectiveness of this DHI on a larger scale.

Conclusions

This process enabled the proposed mechanisms of action and theoretical foundations of a DHI to be comprehensively described, highlighting the key BCTs used to empower patients to self-manage their condition. These findings provide guidance for the ongoing evaluation of the effectiveness of the intervention.

The new standard framework for DHT includes the evaluation of behavior change [10]. There is limited evidence about how people engage with and act upon information delivered via digital behavior change interventions, especially in tools that

span acute to long-term conditions. We have demonstrated that it is possible to evaluate the behavioral mechanisms of this DHT using a range of models and frameworks to illustrate a behavior change model for getUBetter.

Conflicts of Interest

CM is the Clinical Director and Chief Executive Officer of getUBetter.

Multimedia Appendix 1
Examples of APP pages and links to behavior change techniques.
[PPTX File , 3526 KB - formative_v6i3e32669_app1.pptx]

Multimedia Appendix 2
getUBetter—linking the Capability, Opportunity, Motivation–Behavior (COM-B) components to intervention functions and policy categories from the behavior change wheel.
[DOCX File , 24 KB - formative_v6i3e32669_app2.docx]

References

1. The state of musculoskeletal health 2019. Versus Arthritis. 2019. URL: https://www.versusarthritis.org/about-arthritis/data-and-statistics/the-state-of-musculoskeletal-health/ [accessed 2022-02-09]
2. Hutting N, Johnston V, Staal JB, Heerkens YF. Promoting the use of self-management strategies for people with persistent musculoskeletal disorders: the role of physical therapists. J Orthop Sports Phys Ther 2019 Apr;49(4):212-215. [doi: 10.2519/jospt.2019.0605] [Medline: 30931733]
3. Supporting self-management. National Voices. 2014. URL: https://www.nationalvoices.org.uk/publications/our-publications/supporting-self-management [accessed 2022-02-10]
4. de Silva D. Helping people help themselves: a review of the evidence considering whether it is worthwhile to support self-management. The Health Foundation. 2011. URL: https://www.health.org.uk/publications/evidence-helping-people-help-themselves [accessed 2022-02-10]
5. Car J, Koh GC, Foong PS, Wang CJ. Video consultations in primary and specialist care during the COVID-19 pandemic and beyond. Br Med J 2020 Oct 20;371:m3945. [doi: 10.1136/bmj.m3945] [Medline: 33082127]
6. Clarke G, Pariza P, Wolters A. How has COVID-19 affected service delivery in GP practices that offered remote consultations before the pandemic? The Health Foundation. 2020. URL: https://www.health.org.uk/news-and-comment/charts-and-infographics/how-has-covid-19-affected-service-delivery-in-gp-practices [accessed 2022-02-10]
7. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
8. Aalbers T, Baars M, Rikkert MO. Characteristics of effective Internet-mediated interventions to change lifestyle in people aged 50 and older: a systematic review. Ageing Res Rev 2011 Sep;10(4):487-497. [doi: 10.1016/j.arr.2011.05.001] [Medline: 21628005]
9. Walsh N, Pearson J, Healey E. Physiotherapy management of lower limb osteoarthritis. Br Med Bull 2017 Jun 01;122(1):151-161. [doi: 10.1093/bmb/ldx012] [Medline: 28472246]
10. Evidence standards framework for digital health technologies. National Institute for Health and Care Excellence. 2019. URL: https://tinyurl.com/yckphwjy [accessed 2022-02-10]
11. getUBetter. URL: https://www.getubetter.com/ [accessed 2022-02-21]
12. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. Sutton, United Kingdom: Silverback Publishing; 2014:151.
13. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
14. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
15. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24:485-500. [doi: 10.17705/1cais.02428]
16. BCT theory and techniques tool. Human Behaviour Change Project. 2020. URL: https://theoryandtechniquetool.humanbehaviourchange.org/ [accessed 2022-02-10]
17. Atkins L, Francis J, Islam R, O'Connor D, Patey A, Ivers N, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implement Sci 2017 Jun 21;12(1):77 [FREE Full text] [doi: 10.1186/s13012-017-0605-9] [Medline: 28637486]

XSL•FO
RenderX

18. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. Br Med J 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]

19. Hurley DA, Murphy LC, Hayes D, Hall AM, Toomey E, McDonough SM, et al. Using intervention mapping to develop a theory-driven, group-based complex intervention to support self-management of osteoarthritis and low back pain (SOLAS). Implement Sci 2016 Apr 26;11(1):56 [FREE Full text] [doi: 10.1186/s13012-016-0418-2] [Medline: 27113575]

20. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 08;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]

Abbreviations

BCT: behavior change technique
BCTT: Behavior Change Technique Taxonomy
BCTTv1: Behavior Change Technique Taxonomy Version 1
BCW: behavior change wheel
COM–B: Capability, Opportunity, Motivation–Behavior
DHI: digital health intervention
DHT: digital health technology
NHS: National Health Service
NICE: National Institute for Health and Care Excellence
PSD: persuasive systems design

XSL·FO
RenderX

<u>Original Paper</u>

Engagement With Web-Based Fitness Videos on YouTube and Instagram During the COVID-19 Pandemic: Longitudinal Study

Wuyou Sui[1], BSc, MA, PhD; Jonathan Rush[2], PhD; Ryan E Rhodes[3], PhD

[1]Exercise and Health Psychology Lab, Western University, London, ON, Canada

[2]Department of Human Development and Family Studies, College of Health and Human Development, Pennsylvania State University, University Park, PA, United States

[3]Behavioural Medicine Laboratory, School of Exercise Science, Physical and Health Education, University of Victoria, Victoria, BC, Canada

Corresponding Author:
Wuyou Sui, BSc, MA, PhD
Exercise and Health Psychology Lab
Western University
1151 Richmond St
London, ON, N6A 3K7
Canada
Phone: 1 5196612111 ext 81189
Email: yoahsui@gmail.com

Abstract

Background: The COVID-19 pandemic has drastically changed the physical activity (PA) landscape through the closures of gymnasiums, schools, and many outdoor spaces. Physical distancing guidelines have also reduced opportunity for PA. The popularity of free web-based home fitness videos on video hosting platforms (eg, YouTube and Instagram) has spiked during the pandemic. Many web-based fitness videos offer a convenient, accessible, and cost-effective means of engaging in PA through regularly posted videos or discrete programs. Notably, traditional PA programs often suffer from poor adherence and high dropout rates, despite many advantages over web-based workout programs (eg, equipment, feedback, and in-person engagement). Thus, notwithstanding clear advantages of these web-based fitness videos, their ability to maintain long-term engagement and adherence is unknown.

Objective: We explored patterns of engagement (ie, views, likes, and comments) for channels posting daily or program-based web-based fitness videos since the declaration of COVID-19 as a pandemic, over 4 months. Our secondary objective was to examine potential moderators of engagement metrics.

Methods: An environmental scan was used to identify eligible channels. Eligible channels were (1) freely available on YouTube or Instagram and (2) posted daily or weekday series workouts or offered quarantine-specific workout programs. Searches for eligible channels were conducted on June 1 and 4, 2020. Engagement metrics of views, likes, and comments were then collected from channels' videos posted between March 11 and June 26 or 30, 2020, inclusive, on June 26 or July 8, 2020. A series of multilevel modeling analyses were conducted to examine longitudinal changes in each of the 3 outcome variables.

Results: Ten channels were deemed eligible and included in analyses; 6 posted regularly, while the other 4 posted discrete workout programs. Multilevel models revealed that both views and likes significantly decreased across days. Visually, channels display the sharpest drop in engagement within the first week. Linear change estimate indicates that the number of views initially declined by 24,700 per day (95% CI –44,400 to –11,300, P=.01) on average across all the channels. Channels with more subscribers declined in their views, likes, and comments at a significantly higher rate than those with fewer subscribers ($P\leq.04$). The day of the week a video is posted, "virality," and content of a video appear to influence engagement. Integrating behavior change techniques and posting new and varied videos often may help garner further engagement with these videos. Future research should examine common elements of videos, which drive engagement.

Conclusions: Despite raw engagement metrics, each channel demonstrated peak engagement with the initial video followed by decreased engagement with subsequent videos. As many countries maintain restrictions on traditional PA facilities owing to the COVID-19 pandemic, determining methods to improve engagement and adherence with web-based fitness videos becomes increasingly important.

XSL•FO
RenderX

Introduction

The COVID-19 pandemic has scaled up to over 38 million cases and over 1 million deaths worldwide as of this writing [1]. When COVID-19 was declared a pandemic on March 11, 2020 [2], this led to considerable changes in daily life including social distancing, elimination of community gatherings, restricted travel, and general instruction to "stay home" [3]. The consequences have affected how people engage in recreation, as many parks and recreation facilities were closed [4,5]. Commercial gyms, group exercise classes, and sporting activities ceased to operate over safety concerns and many have failed to reopen owing to bankruptcy or continued restrictions [6,7].

As a result of these drastic shifts in opportunity for physical activity (PA), the population has altered daily PA behaviors [8]. Fitbit's blog presentation of its users noted a 5% to 20% reduction in total steps across the globe during the early stages of the pandemic [9]. Early research has replicated this general finding in China [10], the United States [11,12], the United Kingdom [13], Canada [14], and Europe [15]. Furthermore, higher-intensity PA, such as regular exercise, appears to be particularly compromised [12,14,15]. For example, a recent survey of 3052 US adults found a clinically meaningful drop in PA among active participants (pre–COVID-19) who were socially isolating [12]. These findings are concerning because regular moderate to vigorous intensity PA was already low in prevalence among high-income countries before the COVID-19 pandemic [16], despite how essential this behavior is for mental and physical health [17]. An increased burden that this wave of inactivity will cause leaves the population and the health care system weakened and at higher risk as COVID-19 continues to plague many countries [18].

Despite general trends in physical inactivity, some people have reported increasing their PA or remaining stable [14,15]. For example, work by Rhodes et al [14] found that minutes spent in PA during the pandemic were moderately positively associated with personality traits (eg, activity-extraversion, habit, and identity) and weakly positively associated with demographic factors (eg, income, education, and dog ownership). Notably, one of the key predictors of this stability has been opportunities to engage in home exercise [14,19]. Early research examining home exercise delivered through videos or videoconferencing software has demonstrated effectiveness in both improving levels of PA [20,21] and in increasing motivation to be physically active [22]. Further, affordances for PA (equipment and media) in the home environment have always been a reliable correlate of behavior [23]; yet, the COVID-19 pandemic restrictions on recreation facilities and gymnasiums have likely resulted in a dramatic shift in its importance. In concert with this demand, some commercial fitness businesses have transitioned to web-based fitness programs as an alternative to traditional exercise options [24]. Many web-based workout providers are offering quarantine specific workouts in an attempt to serve a new group of viewers, or prior patrons who are looking for a way to maintain their

activity. Web-based fitness programs, on platforms such as YouTube and Instagram channels offer many advantages over traditional fitness options such as cost-effectiveness (free with an internet connection and no or minimal equipment), convenience (autonomy of when to watch and many different options), accessibility (vast catalogue of home fitness workouts and routines suited for different populations, skill levels, modes of exercise, etc), and community (comments, likes or dislikes, and live streaming allow for a level of interactivity that can mimic the social aspects of other traditional workout programs). In fact, YouTube reported a quadrupling of daily views for exercise videos with "no equipment" or "home" in the title since declaration of the pandemic [25].

There are clear advantages of these web-based exercise programs in these unprecedented times; yet, there remain several unanswered questions about long-term effectiveness and engagement. These programs cannot emulate several elements of traditional exercise options, such as movements that require equipment (resistance training machines, bikes, rowing, etc), true in-person feedback, and individualization of training. It may be that dropout from these programs is high. Indeed, fast dropout of approximately 50% from traditional gymnasiums and recreation centers is a well-established phenomenon [26]. More information on the patterns and potential moderators of engagement of these pandemic home exercise workouts will assist in tailoring future workouts toward the more successful approaches (eg, challenges vs daily workouts and population-focused approaches) as well as a general understanding of long-term engagement.

Thus, the purpose of this study was to explore the pattern of engagement levels of 10 web-based freely available exercise channels that posted either daily or program-based fitness videos since the beginning of the COVID-19 pandemic. We explored total engagement and changes in engagement through views, comments, and likes across a period of 4 months during the pandemic. We further explored potential moderators of engagement metrics, such as viewership across a week and engagement across specific video posting formats and video content. Given the exploratory nature of this study, no formal confirmatory hypothesis was proposed. However, based on prior exercise adherence research, we expected to see that engagement with these web-based videos—via higher engagement metrics—would occur for the initial videos posted by channels; however, a sharp drop-off in engagement was also hypothesized to follow in subsequent videos, paralleling retention and adherence patterns in existing PA efforts [26].

Methods

Eligibility Criteria

Eligible media met the following criteria: (1) freely available on YouTube or Instagram and (2) posting daily or weekday series workouts or offering quarantine-specific workout programs, as these criteria were rationalized to parallel closest with regular gymnasium-going or exercise or PA guidelines

[27]. As such, websites that offered free workout series or programs but were not hosted on YouTube or Instagram (eg, CrossFit [28]) were excluded, as comparable metrics of engagement (eg, views, likes, and comments) were not consistently available. Similarly, series or programs that were accessed through a proprietary app (eg, Gymshark conditioning workout app [29]) were also not included. Further, channels that posted regular videos but were not as part of a regular series or program (eg, vlogs, testimonials, and diet advice) were not included, as these unstructured or unrelated videos were thought to receive variable hits depending on their individual popularity (eg, trending or viral videos).

Search Strategy

The following keywords were used (in combination) within search engines: "Daily," "Online," "Workout," "Exercise," "Program," "Streaming," "Fitness," "At-Home," "Of the day," "Today's date" (ie, date of the search), "Class," "Session," "quarantine," "lockdown," and "COVID-19." For example, "Daily online workout program" was a permutation of search strategy used.

Information Sources

Search term keywords and combinations therein were searched directly into a Google search engine, as well as YouTube and Instagram platform search engines. For Google searches specifically, webpages that contained lists of available web-based workouts (eg, "The best places to go for free online workouts" [30]) were parsed for eligible channels.

Procedure

An environmental scan was initially employed to determine the media sources (ie, YouTube and Instagram channels) for longitudinal data extraction. According to Graham [31], environmental scans leverage diverse sources and amounts of data, so decision-makers can be informed of "current social, economic, technological, and political contexts, and to identify any potential short- and long-term shifts." Importantly, the flexibility and diversity of potential data sources within an environmental scan can reveal preliminary trends or current relationships that may warrant further, more robust investigation. As such, two initial searches to identify eligible channels were conducted: one on June 1 and one on June 4, 2020.

Upon selection of appropriate YouTube and Instagram channels, overall engagement metrics were collected for videos posted between March 11 and June 26 or 30, 2020, inclusive. Specifically, these overall engagement metrics were collected on June 26 or July 8, 2020. While this extraction method does infer that older videos have more time to accumulate engagement, the regular or programmed nature of the videos posted by the observed channels is likely still sensitive enough to reflect changes in engagement over time.

Primary Outcome: Engagement Metrics

Engagement metrics from videos of eligible channels were collected; specifically, number of views, number of likes (if available), and number of comments. Further, to explicate trends in video engagement, channels that posted a regular series of workout videos (ie, regularly) were charted differently than channels that released a discrete workout program (ie, program). Further, number of channel subscribers was collected as a potential moderator of channel engagement.

Statistical Analyses

A series of multilevel modeling analyses were conducted to examine longitudinal changes in the engagement metrics. Separate models were carried out for each of the three outcome variables (ie, views, likes, and comments) and were predicted by a linear and quadratic time variable, coded as days since COVID-19 was declared as a global pandemic (March 11, 2020) [2]. Number of channel subscribers (grand-mean centered) and video start date were included as between-channel covariates to statistically adjust for initial differences in channel exposure and time elapsed between the first video and the onset of the pandemic. Subscribers and start date were also included as moderators on both linear and quadratic slopes to determine if the rate of change differed depending on channel size or video onset. All models were estimated in Mplus (version 8.4) using the Bayes estimator to facilitate model convergence. Statistical significance was set at $P<.05$.

Results

Channel Descriptives

Overall, 10 channels were observed in this environmental scan: Fit Factory [32], One Workout A Day [33], The Body Coach [34], Orange Theory Fitness [35], Planet Fitness [36], Holly Dolke [37], Six-Pack Factory [38], Leansquad [39], Amanda Bisk [40], and Gaby Pimental [41]. The majority of these channels were identified through websites found through Google searches that posted curated lists of web-based exercise videos or channels. Descriptive characteristics and statistics for each channel are presented in Tables 1 and 2, respectively. Of the 10 channels observed, 9/10 (90%) were based on YouTube and 1/10 (10%) was based on Instagram. The channels were almost evenly split between regularly posting videos (6/10, 60%) and posting discrete programs (4/10, 40%). Of the channels that posted regularly, post frequency ranged from daily videos (4/6, 67%) to weekdays only (2/6, 33%). Of the channels that posted programs, program length ranged from 7 days (2/4, 50%) to 30 days (1/4, 25%). With respect to posting date, the majority of channels observed began posting in March 2020 (8/10, 80%), with 1 (10%) channel starting in mid-May [32] and 1 (10%) channel starting in June [36]. Subscriber count varied drastically among channels, ranging from 766 [41] to 2,500,000 [34], with a median value of 74,100 (based on the date of data collection).

XSL•FO

RenderX

Table 1. Descriptive characteristics for each channel.

Channel	Channel characteristics			Subscribers, N	Videos posted, N	Video start date
	Platform	Posting format	Target audience			
Fit Factory	Instagram	Regularly (daily)	Adults	30,100	38	May 15, 2020
One Workout a Day	YouTube	Regularly (daily)	Women	77,100	100	March 23, 2020
The Body Coach	YouTube	Regularly (weekdays)	Children and parents	2,500,000	100	March 23, 2020
Orange Theory Fitness	YouTube	Regularly (daily)	Adults	89,300	105	March 18, 2020
Planet Fitness	YouTube	Regularly (weekdays)	Adults and adolescents	71,100	12	June 1, 2020
Holly Dolke	YouTube	Program (30 days)	Women	1,110,000	30	March 23, 2020
Six-Pack Factory	YouTube	Program (7 days)	Adults (men)	1,460,000	7	March 30, 2020
Leansquad	YouTube	Program (7 days)	Adults	21,000	14	March 21, 2020
Amanda Bisk	YouTube	Program (14 days)	Women	24,800	14	March 20, 2020
Gaby Pimental	YouTube	Regularly (daily)	Older adults	766	30	March 18, 2020

Table 2. Descriptive statistics for average video engagement metrics of each channel.

Channel[a]	Views, mean (SD)	Likes, mean (SD)	Comments, mean (SD)
Fit Factory	3396.3 (1679.4)	159.8 (67.4)	17.4 (10.8)
One Workout a Day	4164.4 (8787.1)	209.6 (327.5)	17.0 (15.8)
The Body Coach	1,110,476.1 (1,077,577.6)	23,462.3 (16,401.5)	853.7 (589.8)
Orange Theory Fitness	60,951.4 (37,330.2)	243.8 (244.5)	22.8 (15.9)
Planet Fitness	2652.0 (829.9)	53.3 (29.1)	0.00 (0.00)
Holly Dolke	140,494.2 (120,562.1)	5278.4 (4913.1)	512.0 (421.1)
Six-Pack Factory	9707.3 (7060.8)	299.9 (231.6)	123.7 (63.1)
Leansquad	6811.1 (8272.2)	89.7 (102.9)	8.9 (9.6)
Amanda Bisk	19,175.0 (21,117.2)	266.4 (328.9)	17.6 (10.7)
Gaby Pimental	1071.0 (1337.8)	15.6 (13.3)	6.7 (4.7)

[a]Grand means: 135,889.9 views, 3007.9 likes, and 158 comments.

Engagement Metrics Across the COVID-19 Pandemic

The raw daily engagement metrics are displayed in Figure 1, Multimedia Appendix 1, and Multimedia Appendix 2. Multilevel models were used to formally test for the decline in engagement metrics throughout the pandemic. Our results revealed that both views and likes significantly decreased across days (see linear change estimate in Table 3 and Multimedia Appendix 3, respectively). These linear declines in views and likes were still detectable after adjusting for number of channel subscribers and video start date. Table 3, Multimedia Appendix 3, and Multimedia Appendix 4 also indicate that channels with more subscribers had significantly more views, likes, and comments on average than channels with fewer subscribers. Conversely, channels that began their videos later into the pandemic (ie, had a higher value for their video start date) produced fewer views and likes on average compared to channels that began their videos closer to the onset of the pandemic. Daily views were divided by 10,000 to enable model estimation. Therefore, the intercept value of 82.24 indicates that at the onset of the pandemic (time 0), the estimated number of views would be 822,400. The linear change estimate of –2.47 indicates that the number of views were initially declining by 24,700 per day, on average across all the channels. The positive quadratic change value indicates a slowing of this rapid decline in views. However, the quadratic term was not statistically significantly for any of the engagement metrics. Daily likes and comments were also scaled to enable model estimation (divided by 100 and 10, respectively). Therefore, the initial estimated number of likes would be 12,787 with an initial linear decline of 289 likes per day. Descriptively, visual inspection of Figure 1, Multimedia Appendix 1, and Multimedia Appendix 2 indicate the rapid decline in views, likes, and comments for most of the channels across time. Visual inspection also revealed potential "micro" patterns of engagement. For example, the Orange Theory Fitness channel appears to demonstrate a repeating weekly pattern of engagement, whereby engagement is highest on Monday and dips midweek before rising again during Friday and the weekend. Additionally, uncharacteristic spikes in engagement may indicate the "virality" of some videos.

Figure 1. Trajectories of views for each channel during the COVID-19 pandemic. Values are presented per 1000 subscribers. AB: Amanda Bisk, BC: The Body Coach, FF: Fit Factory, GP: Gaby Pimental, HD: Holly Dolke, LS: Leansquad, OT: Orange Theory Fitness, OW: One Workout a Day, PF: Planet Fitness, SF: Six-Pack Factory.
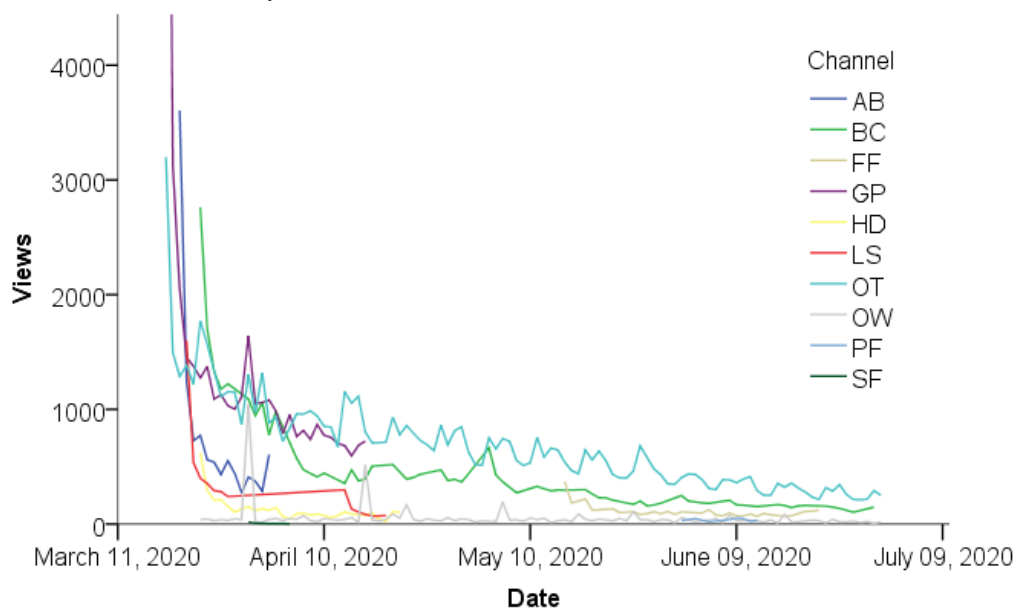
Table 3. Daily changes in views during the COVID-19 pandemic.

Variable	Views[a]		
	Estimate (SE)	95% CI	*P* value
Fixed effects			
Intercept	82.24 (36.13)	2.92 to 145.73	*.04*[b]
Linear change	−2.47 (0.86)	−4.44 to −1.13	*.010*
Quadratic change	0.02 (0.03)	−0.01 to 0.07	.09
Subscribers[c]	0.14 (0.03)	0.07 to 0.21	*.004*
Video start day[d]	−4.35 (3.53)	−12.35 to 1.08	.14
Linear×Subscribers	−0.004 (0.00)	−0.005 to −0.002	*.004*
Linear×Start day	0.11 (0.09)	0.01 to 0.32	*.03*
Quad×Subscribers	0.00 (0.00)	−0.00 to 0.00	.06
Quad×Start day	−0.00 (0.00)	−0.003 to 0.001	.16
Random effects			
Within-person	605.88	525.54 to 692.55	N/A[e]
Between-person			
Intercept	4224.11	1075.70 to 2233.00	N/A
Linear	62.71	0.04 to 1156.84	N/A
Quadratic	0.05	0.01 to 1.30	N/A

[a]Views were divided by 10,000 to enable model estimation.

[b]Italicized values are significant at *P*<.05.

[c]Subscribers = number of channel subscribers / 1000.

[d]Start day = number of days from the beginning of the declaration of COVID-19 as a pandemic (March 11, 2020).

[e]N/A: not applicable.

Moderators of Channel Engagement

The interaction between the number of channel subscribers and linear declines was significant for each of the 3 engagement metrics; that is, channels that had more subscribers experienced a decline in their views, likes, and comments at a significantly faster than those with fewer subscribers. Figure 2, Multimedia Appendix 5, and Multimedia Appendix 6 displays the nature of this interaction for a channel with 500,000 subscribers compared to that with 100,000 subscribers for views, likes, and comments, respectively. Inspection of the figures depicts how channels with more subscribers begin with more engagement but decline at a faster rate than those with fewer subscribers.

Figure 2. Estimated trajectories of views for channels with high (500,000) and low (100,000) subscribers.
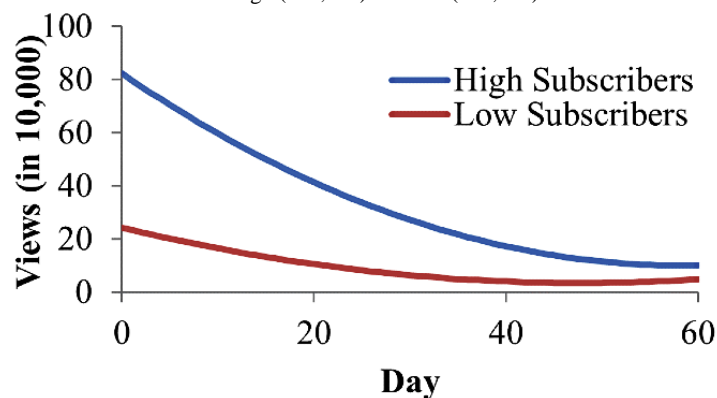
Discussion

Principal Findings

Our study sought to explore the pattern of engagement levels of web-based freely available exercise channels that posted regular or program-based videos since the beginning of the COVID-19 pandemic. To this end, our environmental scan identified 10 channels that met inclusion criteria. While the study is the first of its kind and thus exploratory, we expected to see a large engagement presence for videos posted closer to the start of the pandemic, followed by a sharp drop-off in engagement in subsequent videos, which mirrored patterns among traditional PA opportunities [26]. Overall, commensurate with our expectations, all the channels observed saw a significant linear decline for posted videos over time in views and likes, with a trending decline for comments. Visual inspection indicated the sharpest decline in these engagement metrics within the first week of when the channel started posting videos, controlling for start date, paralleling previously reported dropout rates among traditional gym use [26].

Specifically, earlier videos posted by a channel garnered more views, likes, or comments than subsequent videos, with a drop-off in engagement in subsequent videos, regardless of channel or posting format. The multilevel models also revealed the role of the video start date (in relation to the declaration of COVID-19 as a pandemic), in that channels that began posting videos closer to March 11, 2020, saw higher engagement than those that began posting later. These findings may be representative of the trend in popularity of web-based fitness videos throughout the course of the COVID-19 pandemic. Lockdowns and stay-at-home messaging in early March drastically increased the number of individuals who were staying at home [42], which coincided with a spike in interest in Google searches for "online fitness videos" and "online exercise videos" [43] and a quadrupling of views for "no equipment" and "home" exercise videos on YouTube [25]. However, return to pre–COVID-19 levels of interest in these search terms by May 2020 also generally coincides with plateaus in engagement metrics among videos posted by the observed fitness channels. It is possible that colder or inclement weather may have contributed to the greater engagement in these videos [44] closer to March and a diminishing of engagement over the next few months (ie, winter or spring transition in the Northern hemisphere), in that indoor opportunities for PA—such as web-based workout videos—may have been preferentially chosen over outdoor opportunities in the early months of the COVID-19 pandemic. However, the dramatic drop-off we observed is notably incongruent from the slower moderation of season on PA [44]. Hence, the overall trend we observed supports the notion that the COVID-19 pandemic has driven a mass of engagement to web-based fitness and exercise videos on YouTube and Instagram.

Notably, while our multilevel models revealed a significant negative linear relationship between the video posting date and engagement metrics, when examining our data descriptively, the sharpest decline in all engagement metrics was evident within a week of a channel posting its first video, regardless of the start date. This initial drop-off in channel engagement parallels the poor adherence rates to traditional PA programs, whereby 50% of adults drop out within a few months of beginning a PA program [26]. However, use of YouTube and Instagram as the sole means of delivering a PA program presents several disadvantages to a traditional PA program, which may explain the sharper decrease in engagement. By virtue of the medium of delivery (ie, posting a video on the internet), the web-based fitness videos examined do not incorporate many of the strategies that have been shown to improve adherence to PA. For example, viewers of web-based fitness videos do not inherently have a way to receive feedback on the basis of their participation, which can lead to overexertion or improper form or injury, both of which have been shown to deter future participation in PA [45]. Further, capacity for elements such as goal-setting and social participation with other viewers is limited with web-based fitness videos given the asynchronous "delivery" of the workout.

Descriptively, several potential moderators for engagement were identified within the data. Specifically, while the larger trend of engagement demonstrated was a negative linear relationship, smaller patterns within the data were also observed through visual inspection. For example, some "regularly" posting channels appeared to demonstrate weekly patterns in video views; specifically, video views were higher at the beginning of the week (ie, Monday) and tended to decline throughout the week. This weekly—or circaseptan—cycle has been previously documented for smoking cessation efforts [46] and other health-related behaviors [47], whereby the healthiest contemplations occur during Monday. The Orange Theory Fitness channel [35], in particular, appeared to demonstrate an M-shaped circaseptan cycle, with relatively higher engagement at the start and end of the week. Other notable patterns include large spikes in engagement metrics, which are may be attributable to the "virality" of some videos, or videos that become popular through sharing and resharing [48]. These viral videos do not usually deviate from the creator's typical content, but rather feature some characteristic or characteristics that facilitate traction on social media very soon after publication. Smaller spikes in engagement were also present in regularly posting channels and may be reflective of the specific content of the PA that day. In other words, the specific content of that day's workout may elicit more engagement from viewers than typical or expected.

In general, patterns of engagement metrics were related to other engagement metrics. For example, views of a video were consistently several magnitudes higher than likes on a video; similarly, the number of likes was generally an order of magnitude higher than comments on a video. These findings are unsurprising given that these metrics likely represent an increasing level of engagement of a video, respectively. Views of a video represent a spectrum of viewers: from those who watch the full video (and presumably participate in the PA) to those who may only have a passing curiosity in the video and disengage after 30 seconds of watching the video (ie, minimum watch time to be considered a view [49]). On the other hand, likes and comments are intentionally performed by the viewer, which may be indicative of higher engagement with the video.

Of the channels observed, adults were the primary audience for these videos, followed by women, children and adolescents, and then older adults. Despite insufficient observations to statistically compare channels examining these groups, descriptively, channels demonstrate fairly consistent patterns of engagement. Perhaps a notable exception to this is the channel focused on older adults [41]; Gaby Pimental's channel, despite a subscriber count of <1000, had some of the highest views, likes, and comments per subscriber. Whether this is owing to the low subscriber count (ie, less engagement is needed to achieve higher ratio of engagement) or potentially inherent differences to how older adults engage with these videos (eg, more likely to like and comment compared to other demographics) is worthy of future investigation.

Practical Implications

This work holds several potential practical implications for both future digital exercise research and viewers or creators of digital exercise content. Our data suggest that COVID-19 has generated interest and engagement with web-based fitness and exercise videos. Furthermore, as reports of resurgences of COVID-19 cases emerge globally and many countries revert to lockdowns and stay-at-home orders—and consequently, limitations to traditional PA opportunities—interest in these web-based fitness and exercise videos is likely to rise again. Given the similar drop-off in adherence between the web-based exercise videos observed and traditional PA opportunities (eg, gymnasiums and recreation centers), research investigating means by which these videos can retain engagement with new viewers may be able to draw from existing exercise adherence research; for example, for content creators, posting videos often and regularly, encouraging habit formation [50], posting a variety of exercises [51], responding to viewers' comments as a means of social support and feedback [52], and offering incentives or prizes for regular engagement [53] may all assist content creators in improving and retaining their subscriber base. Similarly, for viewers and subscribers to these channels, goal-setting and creating an action plan on the basis of the posting schedule of a channel may improve adherence [54], and more distally, habit formation [54]. Further, liking and commenting on videos and engaging with the content creator may improve engagement and adherence through fostering social identity [55]. Additionally, exploring avenues by which these videos can be made more accessible or available to digitally disadvantaged populations (eg, low-income households and those with poor digital literacy) is worthy of pursuit, as currently, engagement with these videos—and any benefits imparted from PA derived from these videos—is limited by digital inequities [56].

By extension, research investigating how engagement with these types of videos is linked to actual exercise behavior is also crucial for understanding their effectiveness. Further, building on previous work examining engagement with YouTube videos [57], content analyses examining common elements of more popular exercise videos can reveal potentially unique factors driving engagement in this medium of PA promotion, such as branding, targeting emotions or affective states, and authenticity [58,59]. Similarly, determining demographics for the audiences of these videos can reveal which type of channel or video may be more effective for promoting PA among different populations [58].

Limitations

There are limitations to our work and its interpretation. The primary limitation is that we cannot ascertain the extent to which any individual engagement metric, or combination thereof, translates to actual PA. YouTube only requires 30 seconds of video watch time to be considered a view [49]; similarly, likes and comments—while necessitating more engagement from the viewer—can also be done without participating in any actual exercise. Notably, likes and comments can also only be left by viewers with a YouTube or Instagram account, which considerably limits the use of solely these metrics as a measure of video engagement. Perhaps the best surrogate measure for actual exercise participation may be average watch duration for a video, since it is unlikely for an individual to watch a 30-45–minute workout video without participating in exercise themselves. Unfortunately, these metrics are only available to

the channel owner, along with other useful metrics, such as audience retention (ie, percentage of video watched), number of shares, demographics, and more specific engagement metric trends (eg, changes in day-to-day engagement). Further, some channels posted their videos "live" initially (eg, The Body Coach [34]), and hence have comments and other engagement metrics that are not archived on the webpage. Engagement metrics from these "live" sessions may also provide a better representation of exercise participation; however, these are also only available to the channel owner. Our work also only examined a narrow criterion of freely available channels on YouTube or Instagram. Our searches uncovered numerous more "popular" channels, based on subscriber count and views or uploads, which did not meet our eligibility criteria. Inclusion of these channels through broader inclusion criteria may reveal additional patterns of engagement not observed in this study. Comparably, other platforms (eg, Twitter, Facebook, and paid platforms) may include features to influence engagement beyond what we observed. However, whether metrics for engagement are as easily extracted from these platforms remains an area for future research. Finally, the mode of delivery for these videos (ie, YouTube and Instagram) presents as another limitation, as these videos are only available to individuals with access to these websites. As the COVID-19 pandemic continues to exacerbate digital inequalities [56] (eg, the ability to receive PA promotion and facilitation via web-based exercise videos) the generalizability of these findings is limited to those with access to said videos.

Conclusions

In summary, the declaration of COVID-19 as a global pandemic has coincided with a surge in engagement (ie, views, likes, and comments) with web-based fitness and exercise videos on YouTube and Instagram, with observed videos posted closer to the start of the pandemic garnering significantly more engagement than subsequent videos. However, compared to earlier posted videos, engagement metrics associated with subsequent videos decline significantly, regardless of posting type, with the sharpest decline in engagement occurring within a week of the initial video and greater decline for channels with higher subscriber counts. Investigating means by which these types of channels can improve and maintain engagement with viewers, elements of current popular videos that garner engagement, as well as how this engagement translates to actual exercise behavior, are all lucrative areas for future research, especially throughout the uncertainty associated with the COVID-19 pandemic.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Trajectories of likes for each channel during the COVID-19 pandemic.
[DOCX File , 56 KB - formative_v6i3e25055_app1.docx]

Multimedia Appendix 2
Trajectories of comments for each channel during the COVID-19 pandemic.
[DOCX File , 62 KB - formative_v6i3e25055_app2.docx]

Multimedia Appendix 3
Daily changes in likes during the COVID-19 pandemic.
[DOCX File , 16 KB - formative_v6i3e25055_app3.docx]

Multimedia Appendix 4
Daily changes in comments during the COVID-19 pandemic.
[DOCX File , 15 KB - formative_v6i3e25055_app4.docx]

Multimedia Appendix 5
Estimated trajectories of likes for high (500,000) and low (100,000) subscribers.
[DOCX File , 23 KB - formative_v6i3e25055_app5.docx]

Multimedia Appendix 6
Estimated trajectories of comments for high (500,000) and low (100,000) subscribers.
[DOCX File , 33 KB - formative_v6i3e25055_app6.docx]

References

1. COVID-19 Dashboard. Johns Hopkins University & Medicine. Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-10-14]

2.   Rolling updates on coronavirus disease (COVID-19). World Health Organization. 2020 Jul 31. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen [accessed 2020-10-14]

3.   World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance-publications [accessed 2020-10-14]

4.   Parks Canada calls on visitors to stay home and implements additional measures, suspending visitor vehicle access to Parks Canada places across the country until further notice. Government of Canada. URL: https://www.canada.ca/en/parks-canada/news/2020/03/text.html [accessed 2022-03-03]

5.   NPS Public Health Update. National Park Service. URL: https://www.nps.gov/aboutus/news/public-health-update.htm [accessed 2020-10-14]

6.   Ciison. 2020 Mar 30. URL: https://www.newswire.ca/news-releases/potential-for-massive-closures-amp-bankruptcies-to-fitness-industry-amid-covid-19-pandemic-828115652.html [accessed 2022-03-03]

7.   Murphy C. Gold's Gym permanently closes 30 gyms across the US, due to coronavirus crisis. USA Today. 2020 Apr 16. URL: https://www.usatoday.com/story/money/2020/04/16/covid-19-closures-golds-gym-closes-30-locations-due-coronavirus/5143831002/ [accessed 2020-03-03]

8.   Spence JC, Rhodes RE, McCurdy A, Mangan A, Hopkins D, Mummery WK. Determinants of physical activity among adults in the United Kingdom during the COVID-19 pandemic: The DUK-COVID study. Br J Health Psychol 2021 May;26(2):588-605. [doi: 10.1111/bjhp.12497] [Medline: 33336562]

9.   The Impact Of Coronavirus On Global Activity. Fitbit LLC. 2020 Mar 23. URL: https://blog.fitbit.com/covid-19-global-activity/ [accessed 2022-03-03]

10.   Qin F, Song Y, Nassis GP, Zhao L, Cui S, Lai L, et al. Prevalence of Insufficient Physical Activity, Sedentary Screen Time and Emotional Well-Being During the Early Days of the 2019 Novel Coronavirus (COVID-19) Outbreak in China: A National Cross-Sectional Study. SSRN Journal 2020 Apr 09 [FREE Full text] [doi: 10.2139/ssrn.3566176]

11.   Dunton G, Wang S, Do B, Courtney J. Early Effects of the COVID-19 Pandemic on Physical Activity Locations and Behaviors in Adults Living in the U.S.: Differences by Ethnicity and Income. Cambridge Open Engage 2020 [FREE Full text] [doi: 10.33774/coe-2020-kx2rq-v2]

12.   Meyer J, McDowell C, Lansing J, Brower C, Smith L, Tully M, et al. Erratum: Meyer, J., et al. Changes in Physical Activity and Sedentary Behavior in Response to COVID-19 and Their Associations with Mental Health in 3052 US Adults. 2020, (18), 6469. Int J Environ Res Public Health 2020 Sep 23;17(19):6949 [FREE Full text] [doi: 10.3390/ijerph17196949] [Medline: 32977506]

13.   Rogers N, Waterlow N, Brindle H, Enria L, Eggo RM, Lees S, et al. Behavioral Change Towards Reduced Intensity Physical Activity Is Disproportionately Prevalent Among Adults With Serious Health Issues or Self-Perception of High Risk During the UK COVID-19 Lockdown. Front Public Health 2020;8:575091 [FREE Full text] [doi: 10.3389/fpubh.2020.575091] [Medline: 33102424]

14.   Rhodes RE, Liu S, Lithopoulos A, Zhang C, Garcia-Barrera MA. Correlates of Perceived Physical Activity Transitions during the COVID-19 Pandemic among Canadian Adults. Appl Psychol Health Well Being 2020 Dec;12(4):1157-1182 [FREE Full text] [doi: 10.1111/aphw.12236] [Medline: 33006279]

15.   Cheval B, Sivaramakrishnan H, Maltagliati S, Fessler L, Forestier C, Sarrazin P, et al. Relationships between changes in self-reported physical activity, sedentary behaviour and health during the coronavirus (COVID-19) pandemic in France and Switzerland. J Sports Sci 2021 Mar;39(6):699-704. [doi: 10.1080/02640414.2020.1841396] [Medline: 33118469]

16.   Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Health 2018 Oct;6(10):e1077-e1086 [FREE Full text] [doi: 10.1016/S2214-109X(18)30357-7] [Medline: 30193830]

17.   Rhodes RE, Janssen I, Bredin SS, Warburton DE, Bauman A. Physical activity: Health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug;32(8):942-975. [doi: 10.1080/08870446.2017.1325486] [Medline: 28554222]

18.   Nieman DC. Coronavirus disease-2019: A tocsin to our aging, unfit, corpulent, and immunodeficient society. J Sport Health Sci 2020 Jul;9(4):293-301 [FREE Full text] [doi: 10.1016/j.jshs.2020.05.001] [Medline: 32389882]

19.   Kaushal N, Keith N, Aguiñaga S, Hagger MS. Social Cognition and Socioecological Predictors of Home-Based Physical Activity Intentions, Planning, and Habits during the COVID-19 Pandemic. Behav Sci (Basel) 2020 Aug 31;10(9):133 [FREE Full text] [doi: 10.3390/bs10090133] [Medline: 32878142]

20.   Feng NC, Ryan E, Kidane M, Tusch ES, McFeeley BM, Carlsson R, et al. Feasibility of an at-home, web-based, interactive exercise program for older adults. Alzheimers Dement (N Y) 2019;5:825-833 [FREE Full text] [doi: 10.1016/j.trci.2019.10.005] [Medline: 31799367]

21.   Mascarenhas MN, Chan JM, Vittinghoff E, Van Blarigan EL, Hecht F. Increasing Physical Activity in Mothers Using Video Exercise Groups and Exercise Mobile Apps: Randomized Controlled Trial. J Med Internet Res 2018 May 18;20(5):e179 [FREE Full text] [doi: 10.2196/jmir.9310] [Medline: 29776899]

22.   Shen CC, Yang CC, Mao TY, Sia WY. Do YouTube Fitness Videos Help YouTube user to Learn Fitness? Int J Innov Creat Chang 2019;5(2):93-104 [FREE Full text]

23. Kaushal N, Rhodes RE. The home physical environment and its relationship with physical activity and sedentary behavior: a systematic review. Prev Med 2014 Oct;67:221-237. [doi: 10.1016/j.ypmed.2014.07.026] [Medline: 25084562]

24. Brar F. These Trainers and Studios Are Offering Free Online Workout Classes Amid the Coronavirus Pandemic. Shape. 2020 Apr 28. URL: https://www.shape.com/fitness/trends/free-online-workouts-coronavirus [accessed 2022-03-03]

25. Watching the Pandemic. YouTube Culture & Trends. URL: https://www.youtube.com/trends/articles/covid-impact/ [accessed 2022-03-03]

26. Dishman R. Exercise Adherence: Its Impact on Public Health. J Sport Exerc Psychol 1988;12:202-204 [FREE Full text]

27. Canadian Physical Activity Guidelines for Adults - 18-64 years. Canadian Society for Exercise Physiology. 2011. URL: https://www.healthunit.com/uploads/csep_paguidelines_adults_en.pdf [accessed 2022-03-03]

28. At-Home Workouts. CrossFit. URL: https://www.crossfit.com/at-home/workouts [accessed 2022-03-03]

29. Gymshark. 2021 Mar 30. URL: https://www.gymshark.com/pages/the-gymshark-conditioning-app [accessed 2022-03-03]

30. Oswald E. The best places to go for free online workouts. Digital Trends. 2020 May 01. URL: https://www.digitaltrends.com/health-fitness/best-free-online-workouts/ [accessed 2022-03-03]

31. Graham P, Evitts T, Thomas-MacLean R. Environmental scans: how useful are they for primary care research? Can Fam Physician 2008 Jul;54(7):1022-1023 [FREE Full text] [Medline: 18625830]

32. fitfactoryto. Instagram. URL: https://www.instagram.com/fitfactoryto/?hl=en [accessed 2020-11-26]

33. Life Sessions. YouTube. URL: https://www.youtube.com/channel/UCC7AJ-P3n4j02i-YzReHCNw [accessed 2020-11-26]

34. The Body Coach TV. YouTube. URL: https://www.youtube.com/c/TheBodyCoachTV/featured [accessed 2020-11-26]

35. Orangetheory Fitness. YouTube. URL: https://www.youtube.com/user/Otheoryfitness [accessed 2020-11-26]

36. Planet Fitness. YouTube. URL: https://www.youtube.com/planetfitness/featured [accessed 2020-11-26]

37. Holly Dolke. YouTube. URL: https://www.youtube.com/channel/UCK5PP5Up6Dz80dv5G4XuiiA [accessed 2020-11-26]

38. Sixpackfactory. URL: https://www.sixpackfactory.com/ [accessed 2020-11-26]

39. LEANSQUAD. YouTube. URL: https://www.youtube.com/channel/UCI_VPUgIs60oLH162pVhpOQ [accessed 2020-11-26]

40. Amanda Bisk. YouTube. URL: https://www.youtube.com/channel/UCjPMMZjaZaxOtLKqTni4Qrw [accessed 2020-11-26]

41. Gaby Pimentel's Quaran-tin Workout. YouTube. URL: https://www.youtube.com/c/GabyPimentelsQuarantinWorkout/featured [accessed 2020-11-26]

42. Engle S, Stromme J, Zhou A. Staying at home: The mobility effects of COVID-19. VoxEU. URL: https://voxeu.org/article/staying-home-mobility-effects-covid-19 [accessed 2020-10-14]

43. online. Google Trends. URL: https://trends.google.com/trends/explore?q=online [accessed 2020-10-14]

44. Tucker P, Gilliland J. The effect of season and weather on physical activity: a systematic review. Public Health 2007 Dec;121(12):909-922. [doi: 10.1016/j.puhe.2007.04.009] [Medline: 17920646]

45. Allen K, Morey M. Physical Activity and Adherence. In: Improving Patient Treatment Adherence. New York, NY: Springer; 2010:9-38.

46. Ayers JW, Althouse BM, Johnson M, Cohen JE. Circaseptan (weekly) rhythms in smoking cessation considerations. JAMA Intern Med 2014 Jan;174(1):146-148 [FREE Full text] [doi: 10.1001/jamainternmed.2013.11933] [Medline: 24166181]

47. Ayers JW, Althouse BM, Johnson M, Dredze M, Cohen JE. What's the healthiest day?: Circaseptan (weekly) rhythms in healthy considerations. Am J Prev Med 2014 Jul;47(1):73-76. [doi: 10.1016/j.amepre.2014.02.003] [Medline: 24746375]

48. Broxton T, Interian Y, Vaver J, Wattenhofer M. Catching a viral video. J Intell Inf Syst 2011 Dec 17;40(2):241-259. [doi: 10.1007/s10844-011-0191-2]

49. Baird F. YouTube Analytics: A Simple Guide to Tracking the Right Metrics. Hootsuite. URL: https://blog.hootsuite.com/youtube-analytics/ [accessed 2020-10-14]

50. Rhodes R, Rebar A. Physical Activity Habit: Complexities and Controversies. In: The Psychology of Habit: Theory, Mechanisms, Change, and Contexts. Cham: Springer; 2018:91-109.

51. Sylvester BD, Standage M, McEwan D, Wolf SA, Lubans DR, Eather N, et al. Variety support and exercise adherence behavior: experimental and mediating effects. J Behav Med 2016 Apr;39(2):214-224. [doi: 10.1007/s10865-015-9688-4] [Medline: 26546241]

52. Gomez Quiñonez S, Walthouwer MJL, Schulz DN, de Vries H. mHealth or eHealth? Efficacy, Use, and Appreciation of a Web-Based Computer-Tailored Physical Activity Intervention for Dutch Adults: A Randomized Controlled Trial. J Med Internet Res 2016 Nov 09;18(11):e278 [FREE Full text] [doi: 10.2196/jmir.6171] [Medline: 27829576]

53. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. Am J Prev Med 2007 Oct;33(4):336-345 [FREE Full text] [doi: 10.1016/j.amepre.2007.05.007] [Medline: 17888860]

54. Rhodes R. Chapter Five - The Evolving Understanding of Physical Activity Behavior: A Multi-Process Action Control Approach. Adv Motiv Sci 2017;4:171-205. [doi: 10.1016/bs.adms.2016.11.001]

55. Stevens M, Rees T, Coffee P, Steffens NK, Haslam SA, Polman R. A Social Identity Approach to Understanding and Promoting Physical Activity. Sports Med 2017 Oct;47(10):1911-1918 [FREE Full text] [doi: 10.1007/s40279-017-0720-4] [Medline: 28349449]

56. Beaunoyer E, Dupéré S, Guitton MJ. COVID-19 and digital inequalities: Reciprocal impacts and mitigation strategies. Comput Human Behav 2020 Oct;111:106424 [FREE Full text] [doi: 10.1016/j.chb.2020.106424] [Medline: 32398890]

57. Bärtl M. YouTube channels, uploads and views. Convergence 2018 Jan 10;24(1):16-32. [doi: 10.1177/1354856517736979]

58. Goodyear VA, Armour KM, Wood H. Young people and their engagement with health-related social media: new perspectives. Sport Educ Soc 2019;24(7):673-688 [FREE Full text] [doi: 10.1080/13573322.2017.1423464] [Medline: 31814804]

59. Stollfuß S. Communitainment on Instagram: Fitness Content and Community-Driven Communication as Social Media Entertainment. SAGE Open 2020 May 09;10(2):215824402091953. [doi: 10.1177/2158244020919535]

Abbreviations

PA: physical activity

XSL•FO
RenderX

<u>Original Paper</u>

Smartphone App–Based Noncontact Ecological Momentary Assessment With Experienced and Naïve Older Participants: Feasibility Study

Louise Burke[1*], MSc; Graham Naylor[1*], PhD

Hearing Sciences - Scottish Section, School of Medicine, University of Nottingham, Glasgow, United Kingdom
[*]all authors contributed equally

Corresponding Author:
Graham Naylor, PhD
Hearing Sciences - Scottish Section
School of Medicine
University of Nottingham
New Lister Building Level 3, Glasgow Royal Infirmary
10-16 Alexandra Parade
Glasgow, G31 2ER
United Kingdom
Phone: 44 0141 242 9666
Email: graham.naylor@nottingham.ac.uk

Abstract

Background: Smartphone app–based ecological momentary assessment (EMA) without face-to-face contact between researcher and participant (*app-based noncontact EMA*) potentially provides a valuable data collection tool when geographic, time, and situational factors (eg, COVID-19 restrictions) place constraints on in-person research. Nevertheless, little is known about the feasibility of this method, particularly in older and naïve EMA participants.

Objective: This study aims to assess the feasibility of app-based noncontact EMA as a function of previous EMA experience, by recruiting and comparing a group of participants who had never participated in EMA before against a group of participants who had been part of an earlier in-person EMA study, and age, by recruiting middle-aged to older adults.

Methods: Overall, 151 potential participants were invited via email; 46.4% (70/151) enrolled in the study by completing the baseline questionnaire set and were emailed instructions for the EMA phase. Of these participants, 67% (47/70) downloaded an EMA app and ran the survey sequence for 1 week. In total, 5 daytime surveys and 1 evening survey, each day, assessed participants' listening environment, social activity, and conversational engagement. A semistructured *exit* telephone interview probed the acceptability of the method. As markers of feasibility, we assessed the enrollment rate, study completion rate, reason for noncompletion, EMA survey response rate, and likelihood of reporting an issue with survey alerts and requested assistance from researchers, family, or friends.

Results: Enrollment rates among invitees (63.3% vs 38.2%; P=.004) and completion rates among enrollees (83.9% vs 53.8%; P<.001) were higher in the experienced than in the naïve EMA group. On average, experienced participants responded to 64.1% (SD 30.2%) of the daytime EMA surveys, and naïve participants responded to 54.3% (SD 29.5%) of the daytime EMA surveys (P=.27). Among participants who retrospectively reported issues with survey alerts, only 19% (3/16) requested researcher assistance during data collection. Older participants were more likely to report not being alerted to EMA surveys (P=.008), but age was unrelated to all other markers of feasibility. Post hoc analyses of the effect of the phone operating system on markers of feasibility revealed that response rates were higher among iOS users (mean 74.8%, SD 20.25%) than among Android users (mean 48.5%, SD 31.35%; P=.002).

Conclusions: Smartphone app–based noncontact EMA appears to be feasible, although participants with previous EMA experience, younger participants, and iOS users performed better on certain markers of feasibility. Measures to increase feasibility may include extensive testing of the app with different phone types, encouraging participants to seek timely assistance for any issues experienced, and recruiting participants who have some previous EMA experience where possible. The limitations of this study include participants' varying levels of existing relationship with the researcher and the implications of collecting data during the COVID-19 social restrictions.

XSL•FO
RenderX

Introduction

Background

Ecological momentary assessment (EMA) refers to a range of methods used for measuring daily life feelings, events, experiences, and behaviors in real-time, real-world settings [1]. These methods include pen-and-paper surveys and diary methods [2,3] and surveys delivered via palmtop computers [4,5], mobile phone SMS text messaging [6,7], and, most recently, smartphone apps [8-10]. In the past decade, smartphone app–based EMA has surged in popularity, coinciding with an exponential growth in smartphone ownership; in the United Kingdom, 87% of the population owned a smartphone in 2020 compared with only 27% in 2010 [11,12]. This presents an opportunity for participants to partake in app-based noncontact EMA, where they can download an app and run EMA schedules from their own smartphone, without the need for in-person interaction with researchers at study invitation, recruitment, initiation, or data collection.

Among its advantages, app-based noncontact EMA allows for the recruitment of large samples, or those drawn from specific or hard-to-reach populations, more easily than in-person EMA. It is convenient and can reduce or remove time and geographic barriers to participation. The participant burden associated with borrowing, becoming familiar with, and carrying around a supplied research smartphone is eliminated. For researchers, equipment costs are reduced, and allowing participants to use their own personal smartphone means that less training is required, thus saving time. When training is provided, it is more feasible to use a blanket approach (eg, distribution of generic written instructions only) with noncontact EMA than in-person EMA. Finally, noncontact EMA presents an alternative when situational factors, such as the COVID-19 restrictions, prevent in-person research.

However, there are also limitations associated with this method, not least that participants must own a smartphone. Furthermore, EMA schedules can be complex and demanding, and therefore require a certain degree of participant investment and a solid understanding among participants of what they are being asked to do, along with technological proficiency with the smartphone and app. This may be difficult to achieve without an in-person initiation. Finally, in allowing participants to run EMA on their own smartphones, researchers have less control over the study, and the wide variety of smartphone devices on the market makes it difficult to provide technical support to participants when required. These factors may have negative implications for the quality and quantity of the EMA data collected.

One may reasonably expect that previous experience of successful participation in EMA studies (in whatever form) would increase the likelihood of being able or willing to complete a subsequent app-based noncontact EMA study, although to the authors' knowledge, this has never been explored. In contrast, participants who are new to the EMA method may be disadvantaged by the lack of in-person initiation and subsequently struggle with the noncontact EMA protocol. Regarding the potential effects of age, the high technological demand of EMA may prove challenging for older adults who tend to have lower rates of smartphone ownership, use [13], and competency [14]. Notably, Duncan et al [15] found that, among a sample of sexual minority men, older participants were less willing to download the EMA app to their smartphone than younger individuals. Hence, the limitations of app-based noncontact EMA may lower its feasibility for older participants and introduce sampling bias. As app-based noncontact EMA has become necessary under the COVID-19 restrictions, it is essential to understand the feasibility of this method.

Existing Evidence of Feasibility

To explore the prevalence of published app-based noncontact EMA research and uncover any evidence relating to the abovementioned effects of age and EMA experience, a literature search of the Scopus database was conducted. The search was conducted in April 2020, using the terms *ecological AND momentary AND assessment AND smartphone.* A total of 404 articles were found, all published from 2011 onward. After removing duplicates and inaccessible or irrelevant articles, 73.3% (296/404) of studies remained. Of the 296 studies, 220 (74.3%) used in-person recruitment or initiation techniques, and the studies described by 47 articles were indeterminable as in-person or noncontact EMA. In total, 29 articles reported noncontact procedures, and all were published since 2015. Of these 29 articles, 3 (10%) were protocol papers and 2 (7%) used SMS text message–based sampling measurements. Therefore, 83% (24/29) of publications described observational app-based noncontact EMA studies, demonstrating that this method has been enjoying minimal (in contrast to in-person EMA) but increasing application in recent years.

A meta-analytic approach was used to determine the characteristics of the 24 studies. The median sample size across the studies was 135 participants (range 17-6675). Among studies which reported mean age and dichotomous male–female gender distribution of the sample, participants tended to be younger (mean 30.9, SD 8.0 years, based on 21 studies) and mostly women (mean 68.7% of the sample, based on 19 studies). The topics of study were smoking [16,17], alcohol consumption [18-22], mental health [23-29], dietary behavior [30-35], drug use [36,37], physical activity [38], and tinnitus management [39].

All studies used noncontact methods of recruitment, initiation, and EMA data collection, although the specific procedures varied across studies. Advertisement was conducted on the web via circulation of study details on social media [28,39] and within Reddit and other web-based communities [27,29], and

offline via mass media [26,28], flyers [30], and networking or word of mouth [18,22,30]. When an existing database of university students was available, email invitations were used to advertise studies and recruit samples [21,35,38]. Recruitment and initiation of participants commonly involved directing participants to a webpage where they could read the study description, provide informed consent, complete baseline measures, and access instructions for downloading and using the EMA app [18,33]. In some studies, written instructions for the EMA phase were supplemented by a telephone call [31] or a prerecorded video demonstration [35], indicating that noncontact initiation is more intensive in some studies than others.

Although the specifics of EMA data collection varied depending on the purpose of the study, all 24 studies required participants to run the app on their own smartphone; some specified that this must be running iOS [19] or Android [24], whereas others allowed both [22]. In most studies, the feasibility of this aspect of the design was unaddressed, although McQuoid et al [37] reported that iOS and Android smartphones collected sensor data at different rates in their EMA study; Schlee et al [39] noted low controllability owing to participants using their own smartphones as a limitation of their study, and Wouters et al [35] concluded that the main cause of participant dropout in their study was related to a technical issue caused by an Android update during data collection. Notably, participant dropout rates were not systematically reported; in 71% (17/24) of the studies, it is implied, but not explicitly stated, that no dropout occurred, whereas some reported dropout rates ranging from 14% to 37% [22,23,32-35,38]. Finally, across the 6 studies that did not exclude participants based on low response rate and reported response rate as a percentage of delivered signal-contingent EMA surveys [18-21,32,39], the mean response rate was 65%.

None of the aforementioned 24 studies directly assessed the feasibility of app-based noncontact EMA in comparison with in-person EMA. To the authors' knowledge, the only study to date to have done so is that of Carr et al [40], which was not published when we conducted the literature search. In their sample of men who have sex with men (N=100; mean age 27 years), participants were enrolled in the study and were initiated either in-person or via a live videoconferencing session. The measured outcomes included response rate, behavioral reactivity, and consistency and reliability of the EMA responses. The in-person and noncontact groups returned similar response rates, similar levels of behavioral reactivity, and equally consistent and reliable EMA responses. This suggests that app-based noncontact EMA is feasible and equivalent to EMA initiated in-person, at least with younger, male participants who are initiated via videoconferencing. The use of videoconferencing by Carr et al [40] to initiate noncontact participants means that their findings may not be generalizable to studies that use written instructions [15-17]. Indeed, videoconferencing might arguably be classified as *in-person* for the purposes of instruction and initiation.

Study Objectives

From the foregoing, it is apparent that much remains unknown regarding the feasibility of app-based noncontact EMA.

Specifically, evidence is lacking regarding the method's feasibility with older adults and with experienced versus naïve EMA participants. Therefore, the research questions are as follows:

1. How does the feasibility of app-based noncontact EMA compare, for participants who have never participated in EMA before versus participants who have previous in-person EMA experience?
2. Does older age have an adverse effect on the feasibility of app-based noncontact EMA?

Given the paucity of previous evidence relating to these questions, this study adopted an exploratory approach rather than testing specific hypotheses.

Methods

Participants and Recruitment

In total, 151 members of our pre-existing participant pool were invited. Recruitment occurred in 2 stages. First, 32.5% (49/151) of people who had participated in an earlier in-person EMA study of listening and daily life fatigue conducted by Burke and Naylor [41] (referred to hereafter as the EMA Fatigue study) were invited. In the study, which ran 12 to 18 months before this study, participants were selected randomly from our participant pool, invited by postal letter, and attended three in-person sessions: an initiation session at baseline, a check-in session midway through the study, and a debriefing and interview session at the end. They used smartphones supplied by the researchers with the EMA app preinstalled (the same app as used in this study) to respond to 6 EMA surveys per day for 2 weeks. In all, 50.3% (76/151) of participants initiated the study, and 44.4% (67/151) completed the study. The participants who initiated but subsequently dropped out of that study were not invited to this study to reduce variation with respect to the level of EMA experience acquired. Of the 67 participants who completed the EMA Fatigue study, 18 (27%) were no longer contactable. Thus, 49 individuals with previous EMA experience of 2 weeks, including in-person initiation, were invited to participate in this study. Of these, 63% (31/49) accepted the invitation and formed the *experienced* EMA group.

Thereafter, an additional sample of 102 participants was invited. They had not participated in the EMA Fatigue study, although 19.6% (20/102) were selected randomly and invited to participate in that study—an overlap that resulted from sampling from the same participant pool for both studies. Recruitment of this *naïve* EMA group was conducted on a rolling basis to recruit a group that was similar in size, age, and gender distribution to the experienced EMA group. Initially, 26 invitations were sent, followed by 30 more, followed by 46 more, at which point the naïve group matched the experienced group to an acceptable extent. Of the 39 who accepted the invitation, all but one reported that they had not previously participated in smartphone-based research; one was unsure.

As the aim of recruitment was to recruit as many *experienced* participants as possible from a finite group of past participants, and subsequently to recruit a similarly sized naïve group, this drove our recruitment approach rather than power analyses,

which would have been redundant. Notably, all members of our participant pool were recruited from the National Health Service Audiology in Glasgow, although not all experience hearing loss; issues of tinnitus, hyperacusis, and balance, for example, are also present within our participant pool. The only inclusion criterion was that participants had to own either an Android or iOS smartphone. Invitees' smartphone ownership status was unknown before recruitment.

Ethical Approval

Ethical approval was received from the West of Scotland Research Ethics Committee (18/WS/0007) and the National Health Service Research and Development (GN18EN094). This study was not preregistered.

Procedure

Data were collected from June to August 2020. Email invitations included a participant information sheet and a link to the web-based consent form and baseline questionnaire set. The participants who provided consent and completed the questionnaires were sent a follow-up email containing *beginning the study* and *finishing the study* instruction guides. These participants were asked to download the EMA app to their smartphone and run our EMA survey sequence, which was launched when participants completed a *start-up* survey. At this point, they received an email confirming their enrollment in the study, start and end dates of the EMA phase, and the date and time to expect a telephone *exit* interview. Participants who provided consent and completed the baseline questionnaire set, but did not run the EMA sequence, were emailed a link to a follow-up web-based survey that asked them why they had failed to complete the study.

The EMA sequence consisted of 7 full days of smartphone surveys. Each day, 5 *daytime* surveys were received randomly between 9 AM and 8 PM, with at least an hour between the surveys. Participants had up to 30 minutes to respond to the daytime surveys before they expired. One *evening* survey was received at 9 PM each night, and participants had up to 3 hours to respond. On the final day of the EMA data collection, participants received a *final* survey at 8:30 PM, instead of an evening survey. This informed participants that they had finished the EMA sequence and advised them to consult the *finishing the study* instruction guide, which instructed them to check their data upload status within the app, upload data if necessary, and uninstall the app. The final survey did not expire, so the response window was unlimited.

Telephone *exit* interviews were conducted with all participants who ran the EMA sequence. Participants were made aware of their invitation email that they would be compensated a fixed amount, a £20 (US $27) Amazon e-voucher, to complete the study. They were not paid on a per-survey basis; therefore, there was no monetary incentive to complete the EMA surveys. Participants who completed the baseline survey, 1-week EMA sequence, and telephone interviews received full compensation. It was decided after the data collection, to compensate participants who only completed the baseline survey with a £10 (US $13.5) Amazon e-voucher.

Materials

The web-based consent forms and baseline questionnaires were created and administered using the JISC Online Surveys [42]. The LifeData EMA platform [43] and its corresponding smartphone app, RealLife Exp [44], were used to create and administer smartphone EMA surveys. All the participants used their own smartphones. A participant information sheet explaining the study in detail, a *beginning the study* instruction guide, explaining how to install and use the app and respond to EMA surveys, and a *finishing the study* instruction guide, containing information about data upload and app uninstallation, were supplied as PDF documents.

Measures

Demographic Information

Invitees' age (years), gender (*male, female, or prefer not to say*), degree of hearing loss (represented as better ear average pure-tone threshold across four frequencies: 500 Hz, 1 kHz, 2 kHz, and 4 kHz), and the number of previous in-person appointments attended at our facility were retrieved from the participant database.

Baseline Questionnaires

In all, 3 questionnaires were administered at baseline and are described further. As the study was conducted during the COVID-19 restrictions, some of the questionnaires were modified to better reflect those circumstances. These questionnaires were not directly related to the research questions being examined in this paper; in this context, they merely serve as a procedural step which is often found in EMA studies, namely acquisition of descriptive variables at baseline. As this is a step at which participants may drop out, its presence is relevant for the current purposes. Its content is not relevant, beyond assessing age associations and whether the naïve and experienced groups were equivalent on the variables collected.

An adapted Hearing Handicap Inventory for Adults/the Elderly [45,46] measured the perceived social and emotional consequences of hearing loss. This comprises 26 items and takes approximately 7 minutes to complete. The social activity log [47] was adapted to measure social activity during the past week and month using 15 items and took approximately 5 minutes to complete. Finally, the 10-item Technology Readiness Index (version 2.0 [TRI 2.0]) [48] measured attitudes toward and adoption of technology in the home. A mean value between 1 and 5 was computed for each participant, with higher values reflecting higher levels of techno-readiness. This questionnaire takes approximately 3 minutes to complete. The TRI 2.0 has been found to be both a valid and reliable measurement tool [48].

EMA Surveys

Start-up Survey

A 1-time start-up survey elicited participants' employment status (collapsed into *full- or part-time employed* or *retired/not working/familial caregiver*) and whether they had participated in smartphone-based research previously (*yes*, *no*, or *unsure*).

XSL•FO

RenderX

Daytime Survey

The daytime survey (occurring 5 times per day) elicited self-reports of the type of location, presence or absence of other people, conversational situation, level of background noise, and length of time spent in the situation.

Evening Survey

The evening survey consisted of 6 questions regarding participation in and avoidance of social activity, conversational engagement, and hearing difficulty during that day.

All EMA surveys logged the make, model, and operating system of the phone used to respond.

Follow-up Survey

Participants who completed the baseline questionnaires, but failed to proceed further, were emailed a short JISC survey exploring the reasons for noncompletion of the study. Prefixed by "I did not download the app because," response options were (1) "I did not receive any instructions to do this," (2) "The instructions looked too complicated," (3) "I tried to download the app but found it too difficult/did not know how," (4) "I did not want to," (5) "I did not have time/I was too busy," (6) "I did not realize this was part of the study," (7) "I do not own a smartphone," (8) "My smartphone would not download the app," (9) "I have not got around to it yet," and (10) "Other". Participants were permitted to select only 1 response option.

Exit Interview

The exit interview explored the acceptability of the set-up process (eg, ease of installing the app and completing the start-up survey), study participation (eg, satisfaction with auditory survey notifications and size of question text), procedure for ending the study (eg, data upload status and uninstallation of the app), and general topics (eg, use and usability of instruction guides and reactions of family and friends). In addition, participants with previous EMA experience were asked if they preferred using their own smartphone, as in this study, or a supplied smartphone, as in the EMA Fatigue study.

Data Analysis

Predictor Variables

There were two primary predictors in this study: previous EMA status and age. Participants were classified as experienced EMA participants if they had taken part in the EMA Fatigue study and as naïve EMA participants if they had not. These classifications were corroborated by participant responses to the question which asked if they had previously taken part in any smartphone-based research. Age (years) was treated as a continuous predictor. Interview responses, specifically the high proportion of Android users reporting issues with survey alerts, prompted the post hoc addition of one secondary predictor: phone operating system (coded as *Android* or *iOS*).

Outcome Variables: Markers of Feasibility

The outcome variables were the markers of the feasibility of the app-based noncontact EMA. These were (1) enrollment rate, represented as the percentage of invitees who enrolled into the study by completing the baseline questionnaire set; (2) completion rate, represented as the percentage of enrollees who ran the EMA sequence for 1 week; (3) reason for noncompletion of the study among enrollees, coded dichotomously as *technical reason* or *personal reason*; and (4) EMA survey response rate, measured as the percentage of delivered signal-contingent (ie, daytime and evening) EMA surveys which were responded to. Interview responses prompted the creation of 1 post hoc outcome variable, that being (5) reported issue with survey alerts (coded as *yes* or *no*). Finally, (6) requested assistance from either the research team or a family member or friend, each coded dichotomously as *yes* or *no*, comprised another marker of feasibility. A final outcome measure that applied only to the experienced EMA group was preference for using personal phone or a supplied phone, coded as *preferred own phone*, *preferred supplied phone*, or *no preference*.

Statistical Techniques

First, the associations between the experience group and baseline factors were examined using independent-samples 2-tailed *t* tests and chi-square tests, and the associations between age and baseline factors were assessed using Pearson correlations and binary logistic regression. For all markers of feasibility, summary statistics (means and SDs for continuous variables and frequencies and percentages for categorical variables) were computed for the experienced and naïve groups, and comparisons were made using independent-samples *t* tests and chi-square tests. The associations between age and all markers of feasibility were examined using Pearson correlations and logistic regression analyses. The phone operating system was assessed as a predictor of marker 4 (response rate), and markers 6a and 6b (requested assistance from the research team and from family or friends, respectively), using independent-samples *t* tests and chi-square tests. Finally, responses regarding smartphone preferences were counted for the EMA group.

Across all analyses, chi-square tests were only conducted when the minimum cell count in each group was 5. The analyses were conducted using the SPSS Statistics (version 25; IBM Corp). The α level was set to .05, for all analyses.

Results

Sequence Completion and Markers of Feasibility

Figure 1 illustrates the flow of participants and the number of participants who reached various stages of the study.

Figure 1. Participation at each stage of the study. EMA: ecological momentary assessment.
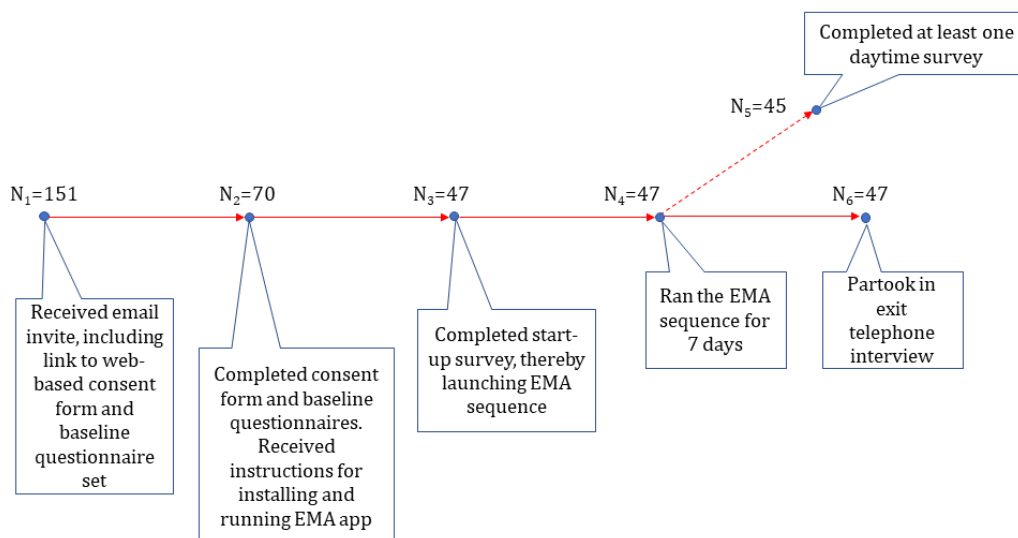
Experience Level

Group Characteristics and Differences at Baseline

The baseline characteristics and comparisons between the experienced and naïve EMA groups are presented in Table 1.

The only baseline variable on which the groups differed was the number of previous laboratory appointments attended, with experienced EMA participants having attended more on average.

Table 1. Baseline characteristics and differences by experience group.

Characteristics	Experienced EMA[a] group (n=26)[b]	Naïve EMA group (n=21)[b]	Group comparisons		
			Chi-square (df)	t test (df)	P value
Age (years), mean (SD), range	66.2 (8.6), 45-78	65.1 (8.6), 46-75	N/A[c]	−0.45 (45)	.65
Gender (male), n (%)	12 (46)	8 (38)	0.31 (1)	N/A	.58
Employed, n (%)	6 (23)	7 (33)	0.61(1)	N/A	.44
Degree of hearing loss in dB HL[d], mean (SD)	22.7 (14.2)	22.1 (13.2)	N/A	−0.14 (45)	.89
Hearing aids users, n (%)	9 (35)	9 (43)	0.33 (1)	N/A	.56
Previous laboratory appointments attended, mean (SD)	5.2 (2.0)	3.3 (2.2)	N/A	−3.04 (45)	.004
HHIA/E[e] score (hearing handicap), mean (SD)	32.3 (29.7)	25.0 (17.6)	N/A	−1.04 (39.88)	.30
SAL[f] score (social activity), mean (SD)	1.9 (1.0)	2.0 (0.9)	N/A	0.34 (44)	.74
TRI[g] 2.0 score (techno-readiness), mean (SD)	3.3 (0.7)	3.2 (0.7)	N/A	−0.14 (44)	.89
Android users, n (%)	15 (58)	12 (57)	0 (1)	N/A	.97

[a]EMA: ecological momentary assessment.

[b]Mean (SD) and t test values are presented for continuous variables and n (%) and chi-square values are presented for categorical variables.

[c]N/A: not applicable.

[d]dB HL: decibels in hearing level.

[e]HHIA/E: Hearing Handicap for Adults/the Elderly.

[f]SAL: social activity log.

[g]TRI: Technology Readiness Index.

Effect of Previous EMA Experience on Markers of Feasibility

As shown in Table 2, the enrollment rate (marker 1) was almost twice as high in the experienced EMA group compared with the naïve EMA group (χ^2_1=8.3; P=.004). Similarly, the completion rate (marker 2) was significantly higher among experienced enrollees compared with their naïve counterparts (χ^2_1=7.1; P=.008). In the experienced group, 60% (3/5) of the participants who enrolled in the study but did not complete it, provided a reason why (marker 3, the only marker not displayed in Table 2); 67% (2/3) cited technical difficulties and 33% (1/3) reported a personal reason. In the naïve EMA group, 44% (8/18) provided a reason why; 25% (2/8) cited technical difficulties and 75% (6/8) cited personal reasons. There were too few responses in each group to analyze group-wise differences in this outcome variable.

Table 2. Markers of feasibility in the overall sample, experienced ecological momentary assessment (EMA) group, and naïve EMA group.

Characteristics	Overall[a]	Experienced[b]	Naïve[c]
N_1 (participants receiving study invitation)	151	49	102
N_2 (participants initiating study)	70	31	39
Marker 1: enrollment rate among invitees (%; N_2/N_1)	46.4	63	38.2
N_3 (participants launching EMA sequence)	47	26	21
N_4 (participants running the 7-day EMA sequence)	47	26	21
Marker 2: completion rate among enrollees (%; N_4/N_2)	67.1	84	53.8
N_5 (participants completing at least one daytime EMA survey)	45	25	20
N_6 (partook in exit interview)	47	26	21
Marker 4: signal-contingent survey response rate (%), mean (SD)	61.4 (30.1)	65.4 (30.7)	56.3 (29.3)
Marker 5: reported survey alert issue (n, yes)	16	7	9
Marker 6a: requested assistance from researcher (n, yes)	3	2	1
Marker 6b: requested assistance from family or friends (n, yes)	6	2	4

[a]Overall sample.

[b]Experienced EMA group.

[c]Naïve EMA group.

The response rate on signal-contingent (ie, daytime and evening) EMA surveys (marker 4) ranged from 0% to 100% in the experienced EMA group and from 0% to 90.9% in the naïve EMA group. The response rate (as shown in Table 2) was not significantly different among the groups (t_{45}=−1.04; P=.31). One participant in each group did not respond to any signal-contingent survey; both reported that they had not been alerted to the incoming surveys. In total, 16 participants reported not being alerted to surveys either some or all the time, as shown in Table 2. There was no effect of previous EMA experience on the likelihood of reporting this issue (χ^2_1=0.0; P=.93). Compared with those who did not report any issue with alerts (marker 5; mean 65.9%, SD 25.6%), those who had attained a lower average response rate (mean 47.7%, SD 34.7%; t_{45}=2.04; P=.047).

In all, 3 participants contacted the researcher for help during the study (marker 6a), all in relation to survey alerts not working, and 6 participants reported asking for help from a family member or friend with some technical aspects of the study (marker 6b). There was too little variation in the responses to assess the effect of past experience on these outcomes.

Age

Relationship of Age to Baseline Variables

Unsurprisingly, older age was related to a greater likelihood of being retired (Wald χ^2_1=10.0; P=.002), more severe hearing loss (r=0.31; P<.001), and a greater likelihood of being a hearing aid user (Wald χ^2_1=5.4; P=.02). Age was also related to gender (Wald χ^2_1=4.5; P=.04), such that older participants were more likely to be men. No association was found between age and the number of previous laboratory appointments attended (r=0.29; P=.05), hearing handicap (r=−0.10; P=.50), social activity (r=0.06; P=.67), techno-readiness (r=0.04; P=.79), or the likelihood of being an Android (vs iPhone) user (Wald χ^2_1=0.0; P=.90).

Effect of Age on Markers of Feasibility

Age was unrelated to both enrollment rate (marker 1; Wald χ^2_1=0.7; P=.39) and completion rate (marker 2; Wald χ^2_1=0; P=.99). The effect of age on the reason for noncompletion (marker 3) was not assessed, as too few responses were obtained. Age was unrelated to response rate (marker 4; r=0.19; P=.21), but older participants were more likely to report not being alerted to surveys (marker 5; Wald χ^2_1=7.0; P=.008). The effect

XSL•FO

RenderX

of age on the likelihood of requesting assistance (markers 6a and 6b) was not examined because of the homogeneity of responses.

Phone Operating System

The phone operating system was related to response rate (marker 4); in comparison with Android users, participants using iOS returned a higher response rate (mean 74.8%, SD 20.3% vs mean 48.5%, SD 31.4%; $t_{45}=-3.28$; $P=.002$). Among iOS users, 25% (5/20) reported not being alerted to surveys (marker 5) compared with 41% (11/27) of Android users; however, this difference was not statistically significant ($\chi^2_1=1.3$; $P=.26$). All 3 participants who sought assistance from the research team (marker 6a) were Android users. Of the 6 who reported asking a family member or friend for help (marker 6b), 4 (67%) were Android users and 2 (33%) were iOS users. The effect of the phone operating system on markers 6a and 6b was not assessed because of the lack of variability of responses on these outcomes.

Although both older age and using an Android operating system were related to reporting issues with survey alerts, the mean age of Android users in this study (mean 65.9, SD 7.9 years) was not significantly different from that of iOS users (mean 65.6, SD 9.4 years; $t_{45}=0.12$; $P=.91$).

Experience of Using Personal Versus Supplied Smartphone

The 26 experienced EMA participants who completed the study were asked if they preferred using their own smartphone, as in this study, or using a supplied smartphone, as in the EMA Fatigue study. Of these 26 participants, 22 (85%) preferred using their own phone, whereas 4 (15%) preferred using the supplied research phone and/or preferred the face-to-face method in general.

Discussion

Principal Findings

This study aimed to assess the feasibility of noncontact EMA. To the authors' knowledge, this is only the second study to do so, after Carr et al [40], and is the first to examine the effects of previous EMA experience, age, and phone operating system on the feasibility of this method. Although the findings suggest that previous in-person EMA experience, younger age, and iOS use may be advantageous, the observed effects were specific to certain markers of feasibility and not systematic across all possible indicators. It is therefore suggested that app-based noncontact EMA is feasible to run with naïve and older EMA participants.

Perhaps the most compelling finding from this study is that naïve participants were less likely to enroll in and complete the study than their experienced counterparts. However, confounding effects could not be ruled out. Specifically, experienced EMA participants had previously attended more in-person laboratory appointments than naïve participants, which may indicate a stronger degree of existing relationship with the researcher or greater willingness to participate in research,

among these participants. Future research would benefit from the measurement and tight control of these factors. The observed higher rate of attrition among naïve participants between the baseline survey and EMA phase is concerning. This may indicate technological difficulties in downloading the app or a lack of understanding of the study demands. A low response to the follow-up survey among dropouts means that little information has been gathered regarding the reasons for attrition; therefore, the feasibility of app-based noncontact EMA in this respect remains unclear. Notably, age was unrelated to both enrollment and completion rates, suggesting that the feasibility of this method is not sensitive to age (within the range studied here; 45-78 years).

Turning to the responses to individual survey alerts, it is notable that the response rate was unrelated to both previous EMA experience and age. At 61.4%, the response rate in this study corresponds closely to the 65% mean of response rates across other noncontact EMA studies [18-21,32,39]. We may have expected a higher response rate, given that this study was conducted at a time when many people were confined to their homes and unable to work, travel, or attend social gatherings owing to the COVID-19 pandemic. Hence, some of the main reasons for nonresponse to EMA surveys, including being in a noisy environment and not hearing the alert, or being in a social situation where it would be inappropriate to respond [41,49], were less likely to occur. However, our participants were not paid per EMA survey completion, as in some other EMA studies [27,29], and it is also possible that participants' motivation to respond about their social interactions was reduced when social activity was restricted. Overall, this study yielded acceptable response rates.

Finally, a sizable number of participants reported that they were not alerted to some or all EMA survey alerts, and their response rates were unsurprisingly lower than those returned by unaffected participants. Whether participants truly experienced a technical issue, or were just unaware of alerts, cannot be determined by the data. However, older participants were disproportionately affected by this issue. These findings undermine the feasibility of app-based noncontact EMA and suggest that supplying participants with a preprogramed smartphone (which would typically involve an in-person initiation session) yields better results in terms of data quantity. However, participants in this study who had experience using both their own smartphone and a supplied research smartphone for EMA overwhelmingly preferred using their own. Future app-based EMA research (both in-person and noncontact) should consider giving participants the option to use either their own smartphone or a supplied smartphone where possible [50], and the chosen app should be tested extensively with different phone makes, models, and operating systems.

Limitations

In addition to caveats mentioned above with respect to specific results, there are several more wide-ranging factors which may limit the reliability and generalizability of this study's results.

First, all participants had working email accounts, possibly indicating some level of technological competence. In support of this, the mean TRI 2.0 score obtained by this sample was

slightly higher than that reported by the Parasuraman and Colby [48] representative sample of US-based adults, reported in the paper in which they introduced the TRI 2.0 (3.33 units vs 3.0 units on a 1-5 Likert scale). Moreover, all participants were recruited from our participant pool, meaning that they had attended at least one in-person appointment at our facility and therefore had some degree of existing relationship with the researchers. This is also a strength of the study, as it indicates some level of homogeneity of participant background, all having come from Glasgow National Health Service Audiology. However, these factors limit the generalizability of our findings. In addition, in the experienced group, the decision not to invite participants who had dropped out of the EMA Fatigue study to recruit a genuinely *experienced* EMA group in this study may have led to the recruitment of an unrepresentatively motivated or conscientious group.

Limited variation and lack of data pertaining to some of the markers of feasibility in this study suggest that they are not useful outcomes. Furthermore, despite attempts to explore the reasons for noncompletion of the study by sending a follow-up survey to the dropouts, only few participants responded. Even less is known about why individuals declined to participate in this study in the first place. Crucially, the smartphone ownership status of this group is not known, a key factor in terms of feasibility.

Finally, our method was noncontact insofar as participants did not attend in-person laboratory sessions; however, they did receive detailed instructions to guide them through the study and were contacted several times by email during recruitment, initiation, and enrollment, as detailed in the *Procedure* section. Consequently, the findings are most relevant to physical noncontact, but nonetheless interactive, EMA research.

Conclusions

This study assessed the feasibility of app-based noncontact EMA as a function of past EMA experience and age. Experienced EMA participants were more likely to enroll in and complete the study than naïve participants, whereas age was unrelated to both enrollment and completion rates. The response rate was acceptable and unrelated to both experience and age, but Android users returned markedly lower response rates than iOS users. Although a sizable number of (mostly older) participants reported in exit interviews that they were not always alerted to surveys, very few informed or sought assistance from the researchers during data collection. In summary, app-based noncontact EMA is feasible, although consideration should be given to how to increase enrollment and completion rates, especially when recruiting participants who are new to EMA.

Conflicts of Interest

None declared.

References

1. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
2. Conner M, Fitter M, Fletcher W. Stress and snacking: a diary study of daily hassles and between-meal snacking. Psychol Health 1999;14(1):51-63. [doi: 10.1080/08870449908407313]
3. Carney MA, Armeli S, Tennen H, Affleck G, O'Neil TP. Positive and negative daily events, perceived stress, and alcohol use: a diary study. J Consult Clin Psychol 2000;68(5):788-798. [Medline: 11068965]
4. Shiffman S. How many cigarettes did you smoke? Assessing cigarette consumption by global report, time-line follow-back, and ecological momentary assessment. Health Psychol 2009;28(5):519-526 [FREE Full text] [doi: 10.1037/a0015197] [Medline: 19751076]
5. Köblitz AR, Magnan RE, McCaul KD, O'Neill HK, Crosby R, Dillard AJ. Smokers' thoughts and worries: a study using ecological momentary assessment. Health Psychol 2009;28(4):484-492 [FREE Full text] [doi: 10.1037/a0014779] [Medline: 19594273]
6. Berkman ET, Dickenson J, Falk EB, Lieberman MD. Using SMS text messaging to assess moderators of smoking reduction: validating a new tool for ecological measurement of health behaviors. Health Psychol 2011;30(2):186-194 [FREE Full text] [doi: 10.1037/a0022201] [Medline: 21401252]
7. Wilson MB, Kallogjeri D, Joplin CN, Gorman MD, Krings JG, Lenze EJ, et al. Ecological momentary assessment of tinnitus using smartphone technology: a pilot study. Otolaryngol Head Neck Surg 2015;152(5):897-903 [FREE Full text] [doi: 10.1177/0194599815569692] [Medline: 25676150]

8. Asselbergs J, Ruwaard J, Ejdys M, Schrader N, Sijbrandij M, Riper H. Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study. J Med Internet Res 2016;18(3):e72 [FREE Full text] [doi: 10.2196/jmir.5505] [Medline: 27025287]

9. Holland E, Koval P, Stratemeyer M, Thomson F, Haslam N. Sexual objectification in women's daily lives: a smartphone ecological momentary assessment study. Br J Soc Psychol 2017;56(2):314-333. [doi: 10.1111/bjso.12152] [Medline: 27484394]

10. Wahl DR, Villinger K, Blumenschein M, König LM, Ziesemer K, Sproesser G, et al. Why we eat what we eat: assessing dispositional and in-the-moment eating motives by using ecological momentary assessment. JMIR Mhealth Uhealth 2020;8(1):e13191 [FREE Full text] [doi: 10.2196/13191] [Medline: 31909719]

11. International communications market report 2010. Ofcom. 2010. URL: https://www.ofcom.org.uk/research-and-data/multi-sector-research/cmr/cmr10 [accessed 2020-10-02]

12. Ofcom national and regions technology tracker - 2020. 9th January to 7th March 2020. Ofcom. 2020. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0037/194878/technology-tracker-2020-uk-data-tables.pdf [accessed 2020-10-02]

13. Mobile consumer survey 2019: the state of the smartphone. Deloitte. 2019. URL: https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/technology-media-telecommunications/deloitte-uk-plateauing-at-the-peak-the-state-of-the-smartphone.pdf [accessed 2021-04-14]

14. Smith A. Older adults and technology use: attitudes, impacts, and barriers to adoption. Pew Research Center. 2014. URL: https://www.pewresearch.org/internet/2014/04/03/attitudes-impacts-and-barriers-to-adoption/ [accessed 2021-04-14]

15. Duncan DT, Park SH, Goedel WC, Sheehan DM, Regan SD, Chaix B. Acceptability of smartphone applications for global positioning system (GPS) and ecological momentary assessment (EMA) research among sexual minority men. PLoS One 2019;14(1):e0210240 [FREE Full text] [doi: 10.1371/journal.pone.0210240] [Medline: 30689651]

16. Roos CR, Brewer JA, O'Malley SS, Garrison KA. Baseline craving strength as a prognostic marker of benefit from smartphone app-based mindfulness training for smoking cessation. Mindfulness 2019;10(10):2165-2171. [doi: 10.1007/s12671-019-01188-6]

17. McQuoid J, Thrul J, Ozer E, Ramo D, Ling PM. Tobacco use in the sexual borderlands: the smoking contexts and practices of bisexual young adults. Health Place 2019;58:102069 [FREE Full text] [doi: 10.1016/j.healthplace.2018.12.010] [Medline: 30639203]

18. Duif M, Thewissen V, Wouters S, Lechner L, Jacobs N. Associations between affect and alcohol consumption in adults: an ecological momentary assessment study. Am J Drug Alcohol Abuse 2020;46(1):88-97. [doi: 10.1080/00952990.2019.1635606] [Medline: 31430201]

19. O'Donnell R, Richardson B, Fuller-Tyszkiewicz M, Liknaitzky P, Arulkadacham L, Dvorak R, et al. Ecological momentary assessment of drinking in young adults: an investigation into social context, affect and motives. Addict Behav 2019;98:106019. [doi: 10.1016/j.addbeh.2019.06.008] [Medline: 31247534]

20. Remmerswaal D, Jongerling J, Jansen PJ, Eielts C, Franken IH. Impaired subjective self-control in alcohol use: an ecological momentary assessment study. Drug Alcohol Depend 2019;204:107479. [doi: 10.1016/j.drugalcdep.2019.04.043] [Medline: 31518888]

21. Slavish DC, Scaglione NM, Hultgren BA, Turrisi R. An ecological momentary assessment of affect, mental health symptoms, and decisions to drink among first-year college women: a pilot study. Prev Sci 2019;20(5):753-764 [FREE Full text] [doi: 10.1007/s11121-018-0966-6] [Medline: 30498934]

22. Poulton A, Pan J, Bruns Jr LR, Sinnott RO, Hester R. Assessment of alcohol intake: retrospective measures versus a smartphone application. Addict Behav 2018;83:35-41. [doi: 10.1016/j.addbeh.2017.11.003] [Medline: 29128148]

23. Levin ME, Haeger J, Cruz RA. Tailoring acceptance and commitment therapy skill coaching in the moment through smartphones: results from a randomized controlled trial. Mindfulness 2018;10(4):689-699. [doi: 10.1007/s12671-018-1004-2]

24. Triantafillou S, Saeb S, Lattie EG, Mohr DC, Kording KP. Relationship between sleep quality and mood: ecological momentary assessment study. JMIR Ment Health 2019;6(3):e12613 [FREE Full text] [doi: 10.2196/12613] [Medline: 30916663]

25. Scherr S, Goering M. Is a self-monitoring app for depression a good place for additional mental health information? Ecological momentary assessment of mental help information seeking among smartphone users. Health Commun 2020;35(8):1004-1012. [doi: 10.1080/10410236.2019.1606135] [Medline: 31025888]

26. von Stumm S. Feeling low, thinking slow? Associations between situational cues, mood and cognitive function. Cogn Emot 2018;32(8):1545-1558. [doi: 10.1080/02699931.2017.1420632] [Medline: 29388502]

27. Mills J, Fuller-Tyszkiewicz M. Nature and consequences of positively-intended fat talk in daily life. Body Image 2018;26:38-49. [doi: 10.1016/j.bodyim.2018.05.004] [Medline: 29883899]

28. Bakolis I, Hammoud R, Smythe M, Gibbons J, Davidson N, Tognin S, et al. Urban mind: using smartphone technologies to investigate the impact of nature on mental well-being in real time. Bioscience 2018;68(2):134-145 [FREE Full text] [doi: 10.1093/biosci/bix149] [Medline: 29599549]

29. Kleiman EM, Turner BJ, Fedor S, Beale EE, Huffman JC, Nock MK. Examination of real-time fluctuations in suicidal ideation and its risk factors: results from two ecological momentary assessment studies. J Abnorm Psychol 2017;126(6):726-738. [doi: 10.1037/abn0000273] [Medline: 28481571]

30. Reichenberger J, Smyth JM, Kuppens P, Blechert J. "I will fast … tomorrow": intentions to restrict eating and actual restriction in daily life and their person-level predictors. Appetite 2019;140:10-18. [doi: 10.1016/j.appet.2019.04.019] [Medline: 31039371]

31. Reichenberger J, Richard A, Smyth JM, Fischer D, Pollatos O, Blechert J. It's craving time: time of day effects on momentary hunger and food craving in daily life. Nutrition 2018;55-56:15-20. [doi: 10.1016/j.nut.2018.03.048] [Medline: 29960151]

32. Wouters S, Thewissen V, Duif M, van Bree RJ, Lechner L, Jacobs N. Habit strength and between-meal snacking in daily life: the moderating role of level of education. Public Health Nutr 2018;21(14):2595-2605 [FREE Full text] [doi: 10.1017/S1368980018001283] [Medline: 29808785]

33. Wouters S, Jacobs N, Duif M, Lechner L, Thewissen V. Affect and between-meal snacking in daily life: the moderating role of gender and age. Psychol Health 2018;33(4):555-572. [doi: 10.1080/08870446.2017.1380813] [Medline: 28934860]

34. Wouters S, Jacobs N, Duif M, Lechner L, Thewissen V. Negative affective stress reactivity: the dampening effect of snacking. Stress Health 2018;34(2):286-295 [FREE Full text] [doi: 10.1002/smi.2788] [Medline: 28971580]

35. Wouters S, Thewissen V, Duif M, Lechner L, Jacobs N. Assessing energy intake in daily life: signal-contingent smartphone application versus event-contingent paper and pencil estimated diet diary. Psychol Belg 2016;56(4):357-369 [FREE Full text] [doi: 10.5334/pb.339] [Medline: 30479445]

36. Sznitman S, Baruch YB, Greene T, Gelkopf M. The association between physical pain and cannabis use in daily life: an experience sampling method. Drug Alcohol Depend 2018;191:294-299. [doi: 10.1016/j.drugalcdep.2018.07.014] [Medline: 30165330]

37. McQuoid J, Thrul J, Ling P. A geographically explicit ecological momentary assessment (GEMA) mixed method for understanding substance use. Soc Sci Med 2018;202:89-98 [FREE Full text] [doi: 10.1016/j.socscimed.2018.02.014] [Medline: 29518701]

38. Marquet O, Alberico C, Hipp AJ. Pokémon GO and physical activity among college students. A study using Ecological Momentary Assessment. Comput Human Behav 2018;81:215-222. [doi: 10.1016/j.chb.2017.12.028]

39. Schlee W, Pryss RC, Probst T, Schobel J, Bachmeier A, Reichert M, et al. Measuring the moment-to-moment variability of tinnitus: the TrackYourTinnitus smart phone app. Front Aging Neurosci 2016;8:294 [FREE Full text] [doi: 10.3389/fnagi.2016.00294] [Medline: 28018210]

40. Carr DJ, Adia AC, Wray TB, Celio MA, Pérez AE, Monti PM. Using the Internet to access key populations in ecological momentary assessment research: comparing adherence, reactivity, and erratic responding across those enrolled remotely versus in-person. Psychol Assess 2020;32(8):768-779 [FREE Full text] [doi: 10.1037/pas0000847] [Medline: 32437190]

41. Burke LA, Naylor G. Daily-life fatigue in mild to moderate hearing impairment: an ecological momentary assessment study. Ear Hear 2020;41(6):1518-1532 [FREE Full text] [doi: 10.1097/AUD.0000000000000888] [Medline: 33136628]

42. Online surveys: powerful, flexible onilne surveys. JISC. 2020. URL: https://www.onlinesurveys.ac.uk/ [accessed 2020-08-31]

43. LifeData software. LifeData. 2020. URL: https://www.lifedatacorp.com/ [accessed 2020-08-31]

44. RealLifeExp mobile app. LifeData. 2015. URL: https://www.lifedatacorp.com/mobile-app/ [accessed 2020-08-31]

45. Ventry IM, Weinstein BE. The hearing handicap inventory for the elderly: a new tool. Ear Hear 1982;3(3):128-134. [doi: 10.1097/00003446-198205000-00006] [Medline: 7095321]

46. Newman CW, Weinstein BE, Jacobson GP, Hug GA. The hearing handicap inventory for adults: psychometric adequacy and audiometric correlates. Ear Hear 1990;11(6):430-433. [doi: 10.1097/00003446-199012000-00004] [Medline: 2073976]

47. Syrjala KL, Stover AC, Yi JC, Artherholt SB, Abrams JR. Measuring social activities and social function in long-term cancer survivors who received hematopoietic stem cell transplantation. Psychooncology 2010;19(5):462-471 [FREE Full text] [doi: 10.1002/pon.1572] [Medline: 19358230]

48. Parasuraman A, Colby CL. An updated and streamlined technology readiness index: TRI 2.0. J Serv Res 2014;18(1):59-74. [doi: 10.1177/1094670514539730]

49. Schinkel-Bielefeld N, Kunz P, Zutz A, Buder B. Evaluation of hearing aids in everyday life using ecological momentary assessment: what situations are we missing? Am J Audiol 2020;29(3S):591-609. [doi: 10.1044/2020_AJA-19-00075] [Medline: 32946256]

50. Ziesemer K, König LM, Boushey CJ, Villinger K, Wahl DR, Butscher S, et al. Occurrence of and reasons for "Missing Events" in mobile dietary assessments: results from three event-based ecological momentary assessment studies. JMIR Mhealth Uhealth 2020;8(10):e15430 [FREE Full text] [doi: 10.2196/15430] [Medline: 33052123]

Abbreviations

EMA: ecological momentary assessment
TRI 2.0: Technology Readiness Index (version 2.0)

XSL·FO

RenderX

XSL•FO
RenderX

Original Paper

Evaluation of a Web-Based Medication Reconciliation Application Within a Primary Care Setting: Cluster-Randomized Controlled Trial

Michael R Gionfriddo[1,2], PharmD, PhD; Yirui Hu[3], PhD; Bhumika Maddineni[2], MPH, PharmD; Melissa Kern[2], MPH; Vanessa Hayduk[2], BSc; William R Kaledas[4], MSc; Nevan Elder[4], BSc; Jeffrey Border[4], MSc; Katie Frusciante[2], MPH; Maria Kobylinski[5], MD; Eric A Wright[2], MPH, PharmD

[1]Division of Pharmaceutical, Administrative and Social Sciences, School of Pharmacy, Duquesne University, Pittsburgh, PA, United States

[2]Center for Pharmacy Innovation and Outcomes, Geisinger Health, Danville, PA, United States

[3]Department of Population Health Sciences, Geisinger Health, Danville, PA, United States

[4]The Steele Institute for Health Innovation, Geisinger Health, Danville, PA, United States

[5]Geisinger Medical Center, Geisinger Health, Danville, PA, United States

Corresponding Author:
Eric A Wright, MPH, PharmD
Center for Pharmacy Innovation and Outcomes
Geisinger Health
100 N Academy Ave Danville
Danville, PA, 17822
United States
Phone: 1 570 714 6633
Email: ewright2@geisinger.edu

Abstract

Background: Despite routine review of medication lists during patient encounters, patients' medication lists are often incomplete and not reflective of actual medication use. Contributing to this situation is the challenge of reconciling medication information from existing health records, along with external locations (eg, pharmacies, other provider/hospital records, and care facilities) and patient-reported use. Advances in the interoperability and digital collection of information provides a foundation for integration of these once disparate information sources.

Objective: We aim to evaluate the effectiveness of and satisfaction with an electronic health record (EHR)-integrated web-based medication reconciliation application, MedTrue (MT).

Methods: We conducted a cluster-randomized controlled trial of MT in 6 primary care clinics within an integrated health care delivery system. Our primary outcome was medication list accuracy, as determined by a pharmacist-collected best-possible medication history (BPMH). Patient and staff perspectives were evaluated through surveys and semistructured interviews.

Results: Overall, 224 patients were recruited and underwent a BPMH with the pharmacist (n=118 [52.7%] usual care [UC], n=106 [47.3%] MT). For our primary outcome of medication list accuracy, 8 (7.5%) patients in the MT arm and 9 (7.6%) in the UC arm had 0 discrepancies (odds ratio=1.01, 95% CI 0.38-2.72, P=.98). The most common discrepancy identified was patients reporting no longer taking a medication (UC mean 2.48 vs MT mean 2.58, P=.21). Patients found MT easy to use and on average would highly recommend MT (average net promoter score=8/10). Staff found MT beneficial but difficult to implement.

Conclusions: The use of a web-based application integrated into the EHR which combines EHR, patient-reported data, and pharmacy-dispensed data did not improve medication list accuracy among a population of primary care patients compared to UC but was well received by patients. Future studies should address the limitations of the current application and assess whether improved implementation strategies would impact the effectiveness of the application.

Introduction

Background

Nearly 50% of Americans are on at least 1 prescription medication [1]. Among Medicare patients, these prescriptions are prescribed on average by 7 different physicians [2]. This does not include nonprescription medications, such as over-the-counter medications, herbal medications, vitamins, and other supplements. The prescribing of medications from multiple providers and the prevalence of nonprescription medication use contribute to inaccurate medication lists. Inaccurate medication lists can lead to adverse consequences, including hospitalization and death [3]. This well-recognized problem has been the subject of initiatives from a variety of organizations [4-7] and is designated a National Patient Safety Goal by the Joint Commission [6].

Medication reconciliation has been recognized as problematic across a variety of health care settings, including primary care. Studies in primary care have found that the rate of discrepancies is high [8-10]. For example, we conducted a study within our primary care environment and identified that when asked, 369 (89.1%) of 414 patients requested a change to their medication list in the electronic health record (EHR) and, on average, patients noted 2.4 discrepancies on their medication lists [8]. At the time of this finding, there were few published tools to facilitate medication reconciliation in primary care and those that did had limitations. For example, a tool developed by Schnipper et al [11] was designed to be used after hospital discharge and compared the discharge medication list to the preadmission list held by the ambulatory care site. This tool was limited in that it was not applicable to all primary care patients and did not facilitate the collection of patient-reported medication use. Conversely, a tool by Lesselroth et al [12] collected medication information from patients using a kiosk; however, this technology was limited to Veterans Affairs facilities and may have missed medications not prescribed outside of those facilities.

To address the limitations of existing tools and improve medication reconciliation within our primary care environment, we developed an EHR-integrated web-based reconciliation application, MedTrue (MT), to facilitate the process of reconciling EHR medication lists with other sources, including patient-reported medications within a primary care environment. We designed this tool to have both staff- and patient-facing interfaces.

Objectives

The goal of this paper is to present the results of an evaluation study aimed at assessing the impact of the application MT on medication list accuracy and patient and staff satisfaction. Using mixed methods, we aimed to evaluate the effectiveness of and satisfaction with MT.

Methods

Study Design

We conducted a mixed methods evaluation of MT, including a pragmatic cluster-randomized controlled trial (cRCT), along with patient and provider surveys and semistructured interviews. This study was approved by the Geisinger Institutional Review Board (2018-0174). Patient use of MT was considered not research as this technology was planned to become part of usual care (UC); however, the evaluative components were considered research.

Setting

The trial was conducted at 6 primary care sites, with an even distribution of rural and urban clinics, within a large integrated health care delivery system. Within our system, there are over 40 primary care sites eligible to participate. The specific sites included were chosen collaboratively with input from primary care and clinic site leadership. Leadership was consulted to ensure alignment with ongoing initiatives and recommended sites based on their capacity to participate in the study (eg, adequate staffing). The included clinics had an average staff size (nurses/physicians/physician assistants) of 10. In the sites, nurses or medical assistants would bring the patient to the room, conduct a medication history and reconciliation, and perform a variety of other tasks to prepare the patient for a visit with their clinician. The clinician would then verify and approve any changes to the patient's medication list. All sites used the same EHR, EpicCare® version 2018 (Madison, WI, USA).

Population

Patients were eligible to participate if they were 18 years of age or older, able to speak English, and seen at a participating site by a member of the primary care team.

Intervention

MT is a real-time web-based application that integrates information from the EHR, patients, and pharmacy-dispensed data into a common database that is viewable to patients through an online portal and in-clinic tablets and to clinicians through an EHR interface. The application was developed using an iterative design process. First, meetings were held with key stakeholders (eg, informatics, clinic staff, and pharmacy) to determine the necessary functionalities of the tool. Based on these initial requirements, we created a minimally viable prototype. We then iteratively improved upon this prototype by implementing it in 2 sites not involved in the later evaluation. These sites would use the tool and provide feedback to the study team. Additionally, study team members would observe the use of the tool in those sites to gather additional insight into how the tool was performing and what adjustments needed to be made. From the feedback received, we created a list of potential improvements, which was prioritized based on necessity and feasibility. The tool tested in this study is the result of several rounds of improvements.

The patient interface presents the patient with their active medication list that is located in the EHR, as well as pharmacy-dispensed data available as an additional function

within the EHR (Figure 1). The patient is asked to review their list of medications, remove medications they are not currently taking, and enter additional medications missing from the list, such as medications purchased over the counter or medications they take that belong to a friend or family member. Additionally, patients are asked about their adherence to their medications over the past month using a Likert-type scale (ranging from 0% for "never" to 100% for "always"). This type of scale is similar to other validated scales for adherence [13,14]. Patients with an upcoming visit at 1 of the intervention sites had access to MT through our online patient portal up to 2 weeks prior to their visit. If the patient did not complete MT through the patient portal prior to their visit, they were provided with a tablet computer (iPad®) by patient access representatives upon check-in and were asked to complete MT while waiting for their appointment. Patients were not given any training in the use of the tool, since during prototyping, we found that most patients were able to navigate the tool without assistance. Approximately a third of the patients accessed MT using the online portal, while the remainder used iPads.

Figure 1. Patient interface.

MT was available to staff through a direct link (access tab) within the EHR as a web-based application. During the intake process and normal workflow of rooming a patient, staff (ie, nurses and medical assistants) accessed MT when assessing medication use. The interface was presented in an embedded browser within the EHR. For the staff interface, MT displayed all the medications the patient was presented with as well as those the patient entered (Figure 2). Discrepancies (ie, patient-suggested removals and additions or pharmacy-dispensed differences) were displayed alongside the EHR-listed medications that were not flagged as discrepant. For each medication adherence, calculated as the proportion of days covered from pharmacy-dispensed data [15], patient-reported medication adherence, captured from the patient-completed Likert-type item, was also displayed. Staff were instructed to conduct their medication history normally and use MT to confirm or change medications. All changes made within MT were "pushed" to the EHR upon clicking the "Save to EHR" button in MT. Staff at the intervention sites had access to paper and audiovisual training materials but also attended an in-person meeting where MT was introduced and reviewed by members of the study team. Study team members were available for consultation and guidance throughout the study.

Figure 2. Staff interface.
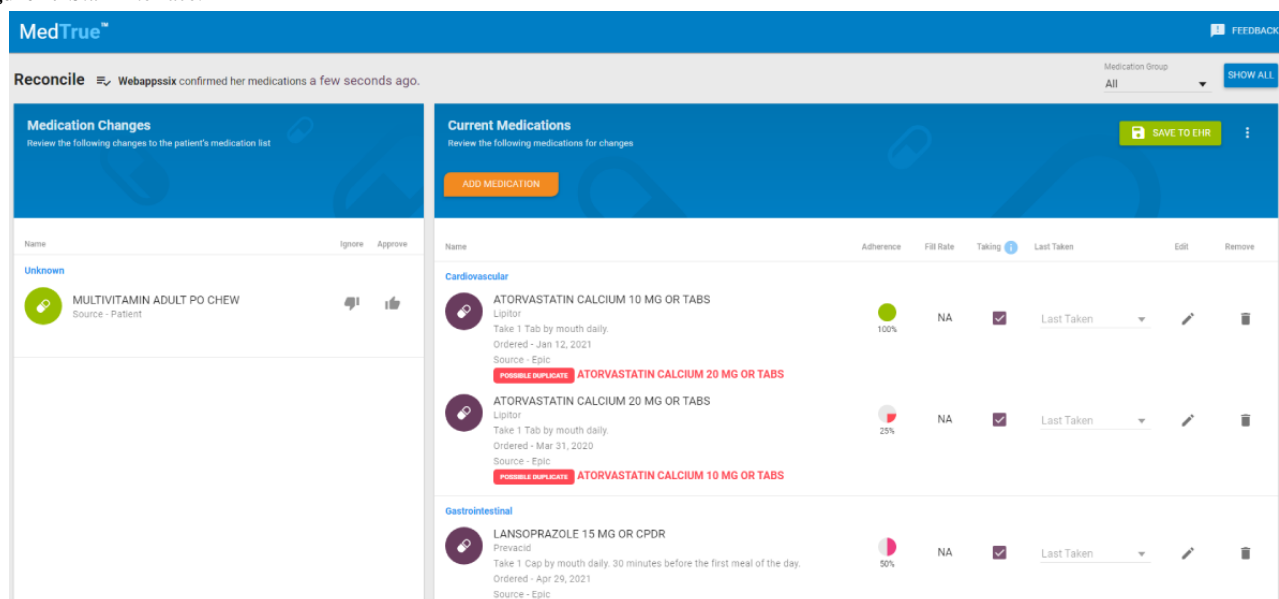
Usual Care

UC providers accessed the EHR-based medication profile that included the standard EHR list of active and discontinued/cancelled medications, along with a separate tab of medications collected through a feed of pharmacy-dispensed data. Patient-reported discrepancies (additions, changes, etc) were also available if the patient completed a previsit questionnaire through the online patient portal (approximately a third of patients have portal accessibility). Standard medication list patient information collected was similar to the MT intervention but did not include self-reported adherence. Hence, UC clinics also had a degree of patient and pharmacy-dispensed information available to clinicians, along with EHR lists, although this information was not available for all patients and information in the EHR was not available within 1 tab. The EHR system used in both UC and MT clinics was EpicCare® version 2018.

Data Collection

Best-Possible Medication History

Trained pharmacists approached a convenience sample of patients exposed to MT as well as a convenience sample of patients in the UC arm to collect a best-possible medication history (BPMH); feasibility considerations informed our decision to use convenience sampling. All pharmacists were licensed and had experience conducting medication histories. Prior to enrolling patients, pharmacists underwent brief training by study staff reviewing components of a BPMH and a clinical interview guide complete with prompts (Multimedia Appendix 1). This guide was based on the literature related to conducting a BPMH [7,16,17]. Pharmacists identified patients via a custom web-based dashboard listing patients with a documented launch of MT by a clinician (MT arm only) or arrival to an eligible visit for UC. Patients completing an encounter in UC were assumed to have EHR-based medication reconciliation due to routine workflow completion of medication reconciliation in all primary care encounters. Pharmacists were blinded to clinic site use of MT, as dashboards only supplied names, not information about MT use. Only study personnel and clinic staff were aware of study group assignment. Pharmacists approached eligible patients after their visit and asked whether they would be interested in participating in a study. If a patient was agreeable, the pharmacist brought them to a private room for consent and to conduct the BPMH.

Pharmacists documented discrepancies for each medication on the postvisit EHR active medication list and also for any medication not present on the medication list. Each medication was coded with only 1 type of medication discrepancy, which fell into 1 of 4 domains. The first was patient acknowledgement of taking the medication differently than prescribed (eg, allergy medication is prescribed daily, but the patient takes it seasonally). Second was removal of a medication on the medication list. The third domain was a modification to the formulation, strength, dose, frequency, route of medication, or timing of administration. The final discrepancy was adding a medication to a profile that was not on the EHR medication list. Since taking a medication differently than prescribed would not automatically infer an inaccurate medication or description of the medication on the medication profile (eg, the patient reported missing a weekend dose of a medication due to forgetfulness), we did not include this in our composite discrepancy endpoint (see later) but did report on this individually. In addition, after each BPMH, pharmacists provided feedback on their confidence that the list they gathered from the patient was accurate using a 3-point scale (not confident, somewhat confident, and confident).

Following the final BPMH collection, study investigators reviewed the reconciled lists of medications in a blinded manner to ensure consistent coding of discrepancies. Inconsistencies were flagged and recoded for final analysis.

Consent, the BPMH, and pharmacist questions were documented using REDCap (Vanderbilt University, Nashville, TN, USA) [18,19]. The interviews took up to an hour, and patients were compensated US $10 for their time in the form of a gift card.

Surveys

All patients exposed to MT were given an opportunity to complete a survey on the usability of and satisfaction with using MT. The survey was presented to the patient following MT completion. Responses were collected within MT.

Staff participating in the intervention arm of the study were also given a survey on the usability of and satisfaction with using MT near the end of the study. Staff were contacted by email and invited to participate in the survey, with up to 2 reminder emails sent at weekly intervals. Surveys were collected using Microsoft Forms (Microsoft Corporation, Redmond, WA, USA).

Interviews

The interviews were guided by a qualitative descriptive approach [20]. Patients in the intervention arm were invited to participate in interviews about their experience using MT through convenience sampling. After a visit where MT was used, patients were approached by a research assistant, who invited them to participate in a brief interview. Consenting patients were interviewed using a semistructured interview guide (Multimedia Appendix 1). The research assistant underwent brief training in qualitative interviewing from 1 of the investigators with experience in qualitative interviewing. The interviews lasted no more than 30 minutes, and patients were compensated US $10 for their time in the form of a gift card.

Rooming staff who used MT during the study period were also given an opportunity to voluntarily participate in semistructured interviews to discuss their experience using MT. The interviews were conducted by an investigator with experience in qualitative interviewing using a semistructured interview guide (Multimedia Appendix 1). The interviews lasted no more than 30 minutes, and staff were not offered compensation.

Primary Outcome

The primary outcome of the study was the accuracy of the medication list. We defined accuracy by counting discrepancies between the medication list gathered through MT or UC and the list gathered by the pharmacist conducting the BPMH. A perfectly accurate medication list would be a list with 0 identified discrepancies.

Secondary Outcomes

We secondarily measured accuracy using a composite numerical assessment of accuracy, where each medication in a list could contribute either 0 or 1 discrepancy, and all discrepancies were totaled per patient. This endpoint included all of the following medication discrepancies: not taking the listed medication, needing the medication to be modified (eg, change in dose or frequency), or having the medication added to their list. Although taking a medication differently was included in our list of discrepancies, this was not included in the composite accuracy list, since this discrepancy would in of itself not necessitate a change in the actual prescription as written, and if the list needed to be modified by the BPMH, 1 of the other listed discrepancies would have been selected over this discrepancy, as per BPMH instructions. We reported individual types of discrepancies identified per group (ie, adding, modifying, removing, or taking a medication differently) on a

medication level. To determine the rate of medication discrepancies per patient, we calculated the percentage of the patients' medications that had discrepancies. Finally, we conducted a sensitivity analysis based on the pharmacists' confidence in their BPMH, including only those patients where the pharmacist conducting the BPMH was confident in having an accurate medication list. Patient and staff satisfaction with MT, as measured by surveys and qualitative interviews, was also reported.

Sample Size

Assuming a 1% intracluster correlation coefficient and a 19% difference in accuracy rates, our target sample size was 300 patients (n=150 [50%] intervention, n=150 [50%] control; each of the 6 sites was expected to enroll 50 [16.7%] participants) to achieve 80% power. During data collection, we conducted an interim look to verify our sample size assumptions and found that our initial assumptions regarding the proportion of completely accurate medication lists (no discrepancies), which were based upon previous studies [21-23], were overly optimistic (~50% in previous studies vs 4% in our interim analysis). As a result, we were underpowered to detect our primary outcome; however, we were powered to detect a secondary composite outcome based on a count of discrepancies. Later, due to feasibility, namely difficulties in recruitment, the trial was stopped.

Randomization

In total, 6 sites were chosen in collaboration with primary care leadership and randomized to the intervention or the control arm using a random number generator in Microsoft Excel (Microsoft Corporation). Prior to recruitment, due to unforeseen challenges, 1 UC site was substituted for a different clinic with similar baseline characteristics.

Blinding

Neither patients nor clinicians were blinded due to the nature of the intervention; however, there were steps taken to blind the pharmacists conducting the BPMH (as described in the Data Collection section). During the analysis, the statistician was blinded as to which arm was intervention and which was control.

Analysis

Qualitative Analysis

Interviews were transcribed verbatim and analyzed thematically [24]. Patient and staff interviews were separately analyzed. The transcripts were independently read by 2 study team members, who collaboratively developed 2 codebooks (1 for patients and 1 for staff). The transcripts were independently coded by 2 study team members using the appropriate codebook. Discrepancies in coding were discussed and resolved by consensus. The codes were then grouped into themes, which were refined based on discussion.

Statistical Analysis

To understand the representativeness of our sample, we compared the demographics of patients seen at the clinics during the intervention period to those who received a BPMH. We described continuous variables using means and SDs as well as

medians and IQRs. Categorical variables were described using frequencies and percentages. Wilcoxon rank-sum tests were used to compare the differences between 2 groups for skewed continuous outcomes; chi-square tests were used to evaluate the independence of categorical outcomes between MT and UC, where odds ratios (ORs) were presented. A mixed effects model with Poisson regression was performed to model the impact of MT on count data (number of medication discrepancies) after adjusting for age, sex, and Charlson comorbidity index, with the total number of medications as the offset variable. Incidence rate ratios (IRRs) and 95% CIs were estimated from Poisson regression models. Sensitivity analyses were performed with or without considering cluster-level covariates for cRCTs. $P < .05$ was considered statistically significant. Statistical analyses were performed using RStudio version 1.2.1335 (RStudio, PBC, Boston, MA, USA) [25].

Additionally, we conducted sensitivity analyses including only those patients for whom the pharmacist rated that they were confident in the history they gathered from the patients; histories where the pharmacist was either somewhat confident or not confident were excluded.

Results

Patient Characteristics

MT was implemented in 3 clinics from July 10, 2018, through August 1, 2019. Over this time frame, MT was accessed 10,835 times by 7342 distinct patients. The majority of access (6501/10,835, 60%) was by tablet computer. Of the 7342 patients using MT, 4005 (54.55%) removed at least 1 medication (mean 3.6 medications removed), 606 (8.25%) added at least 1 medication (mean 1.6 medications added), and 6061 (82.55%) completed the adherence questions. Staff accessed MT for 8443 distinct patients (staff could access MT even if patients did not), removed at least 1 medication for 3544 (41.98%) patients (mean 2.1 medications removed) and added at least 1 medication for 1643 (19.46%) patients (mean 3.3 medications added).

Clinics allocated to MT and UC had similar baseline characteristics, although patients at MT clinics were slightly younger in age (46 vs 48 years), had fewer African Americans (n=871 [3%] vs n=1255 [5%]), less depression (n=1441 [5%] vs n=1490 [6%]), and took more medications (7.75 vs 6.53 medications per patient); see Table 1.

Overall, 224 patients were recruited and underwent a BPMH with the pharmacist (n=106 [47.3%] MT, n=118 [52.7%] UC). The samples of patients who underwent a BPMH were older; were more likely to be female; had depression, dementia, or mild cognitive impairment; were more likely to be married; and had more medications than the baseline clinic population (similar between groups).

Table 1. Demographics.

Characteristic	MT[a] (clinic level), N=30,275	UC[b] (clinic level), N=26,463	*P* value (MT vs UC population)	UC (patient level), N=118	*P* value (population vs sample)	MT patient level), N=106	*P* value (population vs sample)	*P* value (MT vs UC sample)
Age (years), mean (SD)	46 (23)	48 (21)	<.001	55 (18)	<.001	52 (18)	.01	.23
Female, n (%)	15,740 (59.48)	17,886 (59.08)	.35	76 (64.41)	.04	71 (66.98)	.10	.69
Race, n (%)			<.001		.02		.54	.79
Caucasian	29,021 (95.86)	24,882 (94.03)		114 (96.6)		104 (95.3)		
African American	871 (2.88)	1255 (4.74)		3 (2.5)		3 (2.8)		
Other	383 (1.27)	326 (1.23)		1 (0.8)		2 (1.9)		
Ethnicity: Hispanic, n (%)	1243 (4)	1003 (4)	.06	4 (3)	.03	2 (2)	.03	.49
Charlson comorbidity index, mean (SD)	2.20 (2.79)	2.37 (2.84)	<.001	3.12 (2.92)	<.001	2.64 (2.72)	.01	.25
Disorder, n (%)								
MCI or dementia	350 (1.16)	372 (1.41)	.01	5 (4.2)	.01	5 (4.7)	<.001	.86
Depression	1441 (4.76)	1490 (5.63)	<.001	23 (19.49)	<.001	25 (23.58)	<.001	.46
Relationship status (married), n (%)	1682 (5.56)	1478 (5.59)	.88	63 (53.4)	<.001	47 (44.3)	<.001	.18
Number of medications, mean (SD)	7.75 (5.64)	6.53 (5.05)	<.001	12.38 (7.54)	<.001	11.15 (7.29)	<.001	.16

[a]MT: MedTrue.

[b]UC: usual care.

XSL•FO

RenderX

Primary Outcome

For our primary outcome of medication list accuracy (defined as 0 discrepancies), we did not detect a statistically significant difference between MT and UC (OR=1.01, 95% CI 0.38-2.70, P=.98). Pharmacists completing the BPMH reported being confident in the medication history for 200 (89.2%) of 224 patients (91 [85.8%] MT, 109 [92.4%] UC). The primary outcome remained the same upon conducting a sensitivity analysis limited to those BPMHs in which the pharmacist was confident (OR=1.08, 95% CI 0.39-3.02, P=.88).

Secondary Outcomes

The composite secondary outcome of total medication discrepancies, defined as the sum of additions, medications the patient was not taking, and medications that were modified summed per patient, was comparable between MT and UC arms (MT median=4, IQR 2-7 vs UC median=3, IQR 2-6; P=.15). The most common discrepancy was medication the patients' reported not taking, with a median of 2 (IQR 1-3) identified per patient in the UC arm and 1 (IQR 1-3) per patient identified in the MT arm.

The composite outcome was consistent upon sensitivity analysis limited to those patients where a pharmacist was confident in their BPMH (MT median=3, IQR 2-5 vs UC median=4, IQR 2-7; P=.10).

A Poisson regression model indicated that the MT and UC arms were comparable regarding the rate of discrepancies (IRR=1.22, 95% CI 0.97-1.55, P=.09). The rate of discrepancies was significantly affected by age (IRR=1.01, 95% CI 1.00-1.01, P=.03), male sex (IRR=0.81, 95% CI 0.71-0.93, P=.003), and Charlson comorbidity index (IRR=0.95, 95% CI 0.92-0.98, P=.001). Sensitivity analyses with or without considering cluster-level covariates confirmed the robustness of findings of the Poisson regression model (data not shown).

In addition to the composite, we assessed accuracy as the proportion of medications that were discrepant. The proportion of discrepant medications was not significantly affected by MT compared to UC (MT median=40%, IQR 23%-54% vs UC median=34%, IQR 21%-59%; P=.87); this finding was similar on sensitivity analysis (MT median=40%, IQR 22%-55% vs UC median=33%, IQR 21%-56%; P=.66). Further accuracy data can be found in Tables 2-4.

Table 2. Number of patients, with discrepancy type.

Medication discrepancies	UC[a] (N=118), frequency (%)	MT[b] (N=106), frequency (%)	OR[c] (95% CI)	P value[d]
Primary outcome[e]	109 (92.4)	98 (92.5)	1.01 (0.38-2.72)	.98
Modification	72 (61.0)	55 (51.9)	0.69 (0.41-1.17)	.17
Addition	51 (43.2)	42 (39.6)	0.86 (0.51-1.47)	.59
Not taking medication	93 (78.8)	79 (74.5)	0.79 (0.42-1.46)	.45
Taking medications differently	24 (20.3)	22 (20.8)	1.03 (0.54-1.96)	.94

[a]UC: usual care.

[b]MT: MedTrue.

[c]OR: odds ratio.

[d]P value based on chi-square test.

[e]Primary outcome including not taking + modifying + adding medications.

Table 3. Total number of discrepancies per patient.

Medication discrepancies	UC[a] (N=118), mean (SD)	MT[b] (N=106), mean (SD)	UC (N=118), median (IQR)	MT (N=106), median (IQR)	P value[c]
Primary outcome[d]	4.84 (3.85)	4.58 (4.89)	4 (2-7)	3 (2-6)	.15
Modification	1.51 (1.84)	1.29 (1.96)	1 (0-2)	1 (0-2)	.17
Addition	0.85 (1.33)	0.72 (1.23)	0 (0-1)	0 (0-1)	.43
Not taking medication	2.48 (2.60)	2.58 (4.10)	2 (1-3)	1 (0-3)	.21
Taking medications differently	0.33 (0.77)	0.28 (0.61)	0 (0-0)	0 (0-0)	.92

[a]UC: usual care.

[b]MT: MedTrue.

[c]P value based on Wilcoxon test.

[d]Primary outcome including not taking + modifying + adding medications.

XSL•FO
RenderX

Table 4. Percentage of medications on the list discrepant.

Medication discrepancies	MT[a] list %, mean (SD)	UC[b] list %, mean (SD)	MT list %, median (IQR)	UC list %, median (IQR)	P value[c]
Primary outcome[d]	40 (23)	39 (24)	40 (23-54)	34 (21-59)	.87
Modification	11 (13)	10 (14)	8 (0-18)	6 (0-17)	.36
Addition	8 (14)	8 (14)	0 (0-12)	0 (0-11)	.68
Not taking medication	20 (17)	21 (21)	18 (6-29)	16 (1-33)	.82
Taking medications differently	3 (8)	4 (9)	0	0	.63

[a]MT: MedTrue.

[b]UC: usual care.

[c]P value based on Wilcoxon test.

[d]Primary outcome including not taking + modifying + adding medications.

Usability and Satisfaction

Of the 7342 patients who accessed MT, 1450 (19.75% response rate) completed the survey. Patients were satisfied with MT, giving it an average rating of 8 out of 10 for their likelihood to recommend to a friend or family member (Table 5).

Additionally, 1276 (88%) of 1450 patients agreed or strongly agreed that it was easy to use, and 1233 (85%) agreed or strongly agreed that it helped them create a more accurate medication list (Table 6).

Table 5. Patient net promoter score (N=1450).

Patient survey question	Score (1-10)										Mean (SD)
	1, n (%)	2, n (%)	3, n (%)	4, n (%)	5, n (%)	6, n (%)	7, n (%)	8, n (%)	9, n (%)	10, n (%)	
On a scale of 1 (low) to 10 (high), what is the likelihood you would recommend MT[a] to a friend or family member?	41 (2.8)	15 (1)	15 (1)	18 (1.2)	155 (10.7)	112 (7.7)	149 (10.3)	222 (15.3)	156 (10.8)	567 (39.1)	7.95 (2.29)

[a]MT: MedTrue.

Table 6. Patient survey data (N=1450).

Patient survey	Strongly agree, n (%)	Agree, n (%)	Neutral, n (%)	Disagree, n (%)	Strongly disagree, n (%)
I would use MT[a] in the future.	643 (44.3)	568 (39.2)	169 (11.7)	24 (1.7)	46 (3.2)
MT helped me create an accurate medication list.	641 (44.2)	596 (41.1)	124 (8.6)	44 (3.03)	45 (3.1)
MT is easy to use.	724 (49.9)	559 (38.6)	103 (7.1)	17 (1.2)	47 (3.2)
MT reminded me to include all my medications, including those my doctor prescribes and those I purchase over the counter.	721 (49.7)	592 (40.8)	77 (5.3)	21 (1.4)	39 (2.7)

[a]MT: MedTrue.

These findings were echoed in the patients' qualitative interviews and were consistent across sites. Of the 7342 patients who used MT, 14 (0.2%) participated in interviews (1 spouse was also present for 1 of the interviews). Overall, participants had a positive experience with MT. Most found it easy to use and were able to complete the application quickly. The ability to visually see their list of medications rather than be asked about them was noted as an important feature. Seeing the medication names helped patients recognize the medications because they are used to seeing them in print on their medication packaging. Some patients felt it was more efficient than the normal workflow, whereas others disagreed and felt it was creating unnecessary work for them. The patients also noted

that having the indication for the medication would help with recognition and recall. Example quotes from patients illustrating these themes can be found in Table 7.

Of the 22 rooming staff members who accessed MT, 18 (82% response rate) completed the survey. Staff were unsatisfied with MT, giving it an average rating of 0 out of 10 (Table 8).

All staff members disagreed or strongly disagreed that they were satisfied with MT and noted that it did not fit well in their workflow (Table 9).

Interviews provided a more nuanced perspective from staff, with themes and example quotes in Table 10.

Table 7. Patient feedback on MTM[a] (N=14).

MT characteristic	Example quote
Usability	• *. . . It's that simple. I mean, I think somebody from 9 to 90 can use it. I really, I really do, because it's . . . there's nothing confusing about it. It was, and most things are to me, but it truly wasn't. It was quicker than I thought it would be. I thought we'd be there for a while answering questions after. But it wasn't; it was simple.* [Patient T12]
Benefits	• *I think it did help. Because normally, checking in, they would ask what I'm currently taking. This also showed what I was taking 4 months ago but I'm not now. That wouldn't have even crossed my mind if it wasn't right in front of me. I would have never thought to say, "Oh, I'm no longer taking the clobetasol. Oh, I'm not on prednisone anymore." But because it was right in front of me, I thought, oh yeah, I can remove that. I can remove that. I found that to be helpful.* [Patient T13] • *I thought it was easy, and I like that I can delete things off of there if I want. Because I've been asking for 1 medication to be taken off and it hadn't, so I was able to do that.* [Patient P7] • *It's just time-saving. I mean, you're going to sit there and wait for the doctor anyway, so you might as well get it done. And then, it makes it quicker once you're back in the room. You have control over what's in your chart—which I feel like, sometimes, you don't, and, you know, you go to another doctor, and they're like, "Oh, I see you're taking this." And it's like, I haven't taken that in a million years, you know?* [Patient P7]
Necessity	• *. . . It's just an extra thing to do on our part, and I really wonder if it's really necessary to get that information.* [Patient T9]
Suggestions	• *On the meds, the listings of the meds, you'd had both the marketing name and you have the commonly known. The other thing, if you could do it simple, after that say, you know, blood thinner or cholesterol drug for the average person. That's what they think about it. "Oh, that's my blood pressure med. That's my cholesterol med." Because if not, they look at it, you know, hydrochlorothiazide. They're like, "I don't know what that is."* [Patient T8]

[a]MT: MedTrue.

Table 8. Staff net promoter score (N=18).

Staff survey question	Score (1-10)										Mean (SD)
	1, n (%)	2, n (%)	3, n (%)	4, n (%)	5, n (%)	6, n (%)	7, n (%)	8, n (%)	9, n (%)	10, n (%)	
On a scale of 1 (low) to 10 (high), what is the likelihood you would recommend MT[a] to a friend or family member?	16 (88)	1 (6)	—[b]	—	1 (6)	—	—	—	—	—	0.28 (0.96)

[a]MT: MedTrue.

[b]Not applicable.

Table 9. Staff survey data (N=18).

Staff survey	Strongly agree, n (%)	Agree, n (%)	Neutral, n (%)	Disagree, n (%)	Strongly disagree, n (%)
I am satisfied with MT[a].	—[b]	—	—	4 (22)	14 (78)
I want to use MT in the future.	—	—	—	3 (17)	15 (83)
I would want to use MT at every visit with every patient.	—	—	—	3 (17)	15 (83)
MT helped me conduct medication reconciliation with my patients.	—	—	—	6 (33)	12 (67)
MT is compatible with my clinic flow.	—	—	—	3 (17)	15 (83)
MT is easy to use.	—	1 (6)	—	3 (17)	14 (78)

[a]MT: MedTrue.

[b]Not applicable.

Table 10. Staff feedback on MT[a] (N=11).

MT characteristic	Example quote
Usability	• *Yeah, like overall, if it wasn't so slow and it didn't refresh after every time you clicked something, there really would've been, like, no, like, huge issue with it; like, it was just the fact that, like, every time you clicked a box, it just spun—which, like, you know, really held you up.* [Nurse P4] • *Oh, it was very usable. It was very easy. Like again minus the slow Wi-Fi once we got it up and running it was like very easy to get through and you know kind of clear, to the point, you can click on things very easily.* [Nurse P1]
Benefits	• *I've been able to say, well, you know, you say you're taking your Lasix, but it looks like you've only been getting it filled, like, 60% of the time. Is there a reason, you know, that you weren't getting it filled? Because I'm concerned that maybe you aren't taking it like you should. So, it opened up that dialogue . . .* [Nurse MT9] • *I think the benefits of MT are we get a more accurate picture of the medications patients are on, which in the long run helps us treat them better.* [Nurse MT9]
Functionality	• *MT was not able to do that. Um, in [the medical record], they have, like, a little Post-It. It looks like a Post-It note next to the medication, and you just click on that, and you can enter in a comment. And I can say [the] patient has not started yet or will start or you know was too expensive.* [Nurse P1]
Patient Concerns	• *They will remove something in the waiting room, and then they get frustrated that we're still going over it with them in the room.* [Nurse MT7]

[a]MT: MedTrue.

Discussion

Principal Results

The use of MT, a web-based application facilitating medication reconciliation, did not impact the accuracy of medication lists. The application was well received by patients; however, implementation challenges limited usability and acceptance by staff. The failure to improve medication list accuracy reflects both implementation challenges as well as the challenge of reliance on a technology-only solution to obtaining accurate medication lists.

Comparison With Prior Work

Our application is 1 of many technologies and applications being developed to obtain and maintain accurate medication lists [26-32]. Many applications are focused on hospitalized patients, and not all applications have been pragmatically evaluated or included a methodology to compare accuracy with the gold-standard BPMH. Lesselroth et al [12,33] have developed a medication reconciliation kiosk where patients can review and edit their medication lists. Similar to MT, this information is imported into the EHR and reviewed by clinicians [31]. In their evaluation, they found similar rates of discrepancies (29%-39% vs 34%-40%), and the overall number of patients with discrepancies were similar (91%-99% vs 92%-93%) [33,34]. When evaluating the implementation of their application, they also found that time is a factor in using medication reconciliation technology and that, to be taken up, the technology needs to be well integrated into current workflows [35]. This was also found to be a challenge with our application despite its intent to work within the staff workflow.

Strengths

We implemented several design elements to enhance the validity of our evaluation. We randomized clinics to MT or UC to avoid potential contamination that might have occurred at the patient or clinician level. We controlled for possible confounding factors through an adjusted analysis, the results of which mirrored the unadjusted analysis, increasing our confidence in our results. Throughout the trial, we also attempted blinding, wherever possible, to minimize bias. Other strengths include the use of pharmacists to conduct a BPMH and our application itself, which incorporates patient, clinician, and pharmacy-dispensed data to facilitate medication reconciliation.

Limitations

Despite these strengths, a few limitations should be noted. Due to the nature of the intervention, clinicians and, to an extent, patients could not be blinded to the intervention. Fidelity to the use of MT was not always optimal and was variably used among rooming staff. As indicated in the staff survey and interviews, MT was not favorably viewed by the staff and as such they might have failed to use MT consistently, which would reduce observed effectiveness. Technical challenges, such as loading delays and internet connectivity for tablets, also affected the usability of the application in our clinics. Additionally, although the application was implemented within the EHR, it was outside of the existing workflow and required several additional clicks. Finally, we are underpowered to confidently detect the differences we observed in this trial. This was, in part, due to optimistic initial estimates of the effect of MT, as well as recruitment challenges. Our initial estimates of the effect of MT may have been optimistic as by the time the application was implemented, certain features were also available in the EHR, thereby reducing the difference between MT and UC.

Conclusion

Novel approaches to medication reconciliation are needed to ensure safe and effective medication use. The use of technology, such as MT, a web-based application integrating various data sources, is a promising solution to reduce the rate of

discrepancies. In our study, MT did not affect the rate of discrepancies, which may have been the result of failure to use the tool consistently and with fidelity due to implementation challenges. Thus, technology solutions may need to be paired with additional implementation strategies, and particular attention should be paid to how these technologies and other innovations are integrated into clinical workflows to facilitate fidelity and maximize effectiveness.

Acknowledgments

Conflicts of Interest

None declared.

Editorial notice: This randomized study was not registered. The authors explained that the trial was intended to be a part of a staged roll-out within usual care to improve the quality of treatment received by patients. Randomization was used to assign which sites received the improvements first. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials, because the risk of bias appears low and the study was considered formative. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness.

Multimedia Appendix 1
Qualitative interview guides.
[DOCX File , 19 KB - formative_v6i3e33488_app1.docx]

Multimedia Appendix 2
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1309 KB - formative_v6i3e33488_app2.pdf]

References

1. Gindi R, Holmes JS, Arispe IE. Health, United States 2017: With Special Feature on Mortality. 2018. URL: https://stacks.cdc.gov/view/cdc/59199 [accessed 2022-02-17]
2. Pham HH, Schrag D, O'Malley AS, Wu B, Bach PB. Care patterns in Medicare and their implications for pay for performance. N Engl J Med 2007 Mar 15;356(11):1130-1139. [doi: 10.1056/nejmsa063979]
3. Santell JP. Reconciliation failures lead to medication errors. Jt Comm J Qual Patient Saf 2006 Apr;32(4):225-229. [doi: 10.1016/s1553-7250(06)32029-6]
4. Institute for Healthcare Improvement. Medication Reconciliation to Prevent Adverse Drug Events. 2022. URL: http://www.ihi.org/Topics/ADEsMedicationReconciliation/Pages/default.aspx [accessed 2022-02-17]
5. Society of Hospital Medicine. Medication Reconciliation for Hospitalists. 2022. URL: https://www.hospitalmedicine.org/clinical-topics/medication-reconciliation/ [accessed 2022-02-17]
6. The Joint Commission. National Patient Safety Goals. 2022. URL: https://www.jointcommission.org/standards_information/npsgs.aspx [accessed 2022-02-17]
7. World Health Organization. High 5s: Standard Operating Procedures. 2022. URL: https://www.who.int/patientsafety/topics/high-5s/en/ [accessed 2022-02-17]
8. Dullabh P, Sondheimer N, Katsh E, Evans MA. How patients can improve the accuracy of their medical records. EGEMS (Wash DC) 2014 Jan 01;2(3):1080 [FREE Full text] [doi: 10.13063/2327-9214.1080] [Medline: 25848614]
9. Andrus MR. Student pharmacist initiated medication reconciliation in the outpatient setting. Pharm Pract (Granada) 2012 Apr;10(2):78-82 [FREE Full text] [doi: 10.4321/s1886-36552012000200003] [Medline: 24155821]
10. Varkey P, Cunningham J, Bisping S. Improving medication reconciliation in the outpatient setting. Jt Comm J Qual Patient Saf 2007 May;33(5):286-292. [doi: 10.1016/s1553-7250(07)33033-x]
11. Schnipper J, Liang C, Hamann C, Karson A, Palchuk M, McCarthy P, et al. Development of a tool within the electronic medical record to facilitate medication reconciliation after hospital discharge. J Am Med Inform Assoc 2011 May 01;18(3):309-313 [FREE Full text] [doi: 10.1136/amiajnl-2010-000040] [Medline: 21486889]
12. Lesselroth BJ, Felder RS, Adams SM, Cauthers PD, Dorr DA, Wong GJ, et al. Design and implementation of a medication reconciliation kiosk: the Automated Patient History Intake Device (APHID). J Am Med Inform Assoc 2009 May 01;16(3):300-304. [doi: 10.1197/jamia.m2642]
13. Feldman BJ, Fredericksen RJ, Crane PK, Safren SA, Mugavero MJ, Willig JH, et al. Evaluation of the single-item self-rating adherence scale for use in routine clinical care of people living with HIV. AIDS Behav 2013 Jan 30;17(1):307-318 [FREE Full text] [doi: 10.1007/s10461-012-0326-7] [Medline: 23108721]

XSL·FO
RenderX

14. Lu M, Safren SA, Skolnik PR, Rogers WH, Coady W, Hardy H, et al. Optimal recall period and response task for self-reported HIV medication adherence. AIDS Behav 2008 Jan 19;12(1):86-94. [doi: 10.1007/s10461-007-9261-4] [Medline: 17577653]

15. Nau DP. Proportion of days covered (PDC) as a preferred method of measuring medication adherence. Pharmacy Quality Alliance. 2011. URL: http://ep.yimg.com/ty/cdn/epill/pdcmpr.pdf [accessed 2022-02-17]

16. Drenth-van Maanen AC, Spee J, van Hensbergen L, Jansen PAF, Egberts TCG, van Marum RJ. Structured history taking of medication use reveals iatrogenic harm due to discrepancies in medication histories in hospital and pharmacy records. J Am Geriatr Soc 2011 Oct 18;59(10):1976-1977. [doi: 10.1111/j.1532-5415.2011.03610_11.x] [Medline: 22091519]

17. MARQUIS Investigators. MARQUIS Implementation Manual: A guide for medication reconciliation quality improvement. Society of Hospital Medicine. 2014. URL: https://www.hospitalmedicine.org/globalassets/clinical-topics/clinical-pdf/shm_medication_reconciliation_guide.pdf [accessed 2022-02-17]

18. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]

19. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap): a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]

20. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g]

21. Wolff CM, Nowacki AS, Yeh J, Hickner JM. A randomized controlled trial of two interventions to improve medication reconciliation. J Am Board Fam Med 2014 May 07;27(3):347-355 [FREE Full text] [doi: 10.3122/jabfm.2014.03.130240] [Medline: 24808113]

22. Tam VC, Knowles SR, Cornish PL, Fine N, Marchesano R, Etchells EE. Frequency, type and clinical importance of medication history errors at admission to hospital: a systematic review. CMAJ 2005 Aug 30;173(5):510-515 [FREE Full text] [doi: 10.1503/cmaj.045311] [Medline: 16129874]

23. Leonhardt KK, Pagel P, Bonin D, Moberg DP, Dvorak ML, Hatlie MJ. Creating an accurate medication list in the outpatient setting through a patient-centered approach. In: Henriksen K, Battles JB, Keyes MA, Grady ML, editors. Advances in Patient Safety: New Directions and Alternative Approaches. Rockville, MD: Agency for Healthcare Research and Quality; Aug 2008.

24. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

25. Rizopoulos D. GLMMadaptive: Generalized Linear Mixed Models using Adaptive Gaussian Quadrature, R package version 0.5-1. 2022. URL: https://drizopoulos.github.io/GLMMadaptive/ [accessed 2022-02-17]

26. Kramer H, Gibson B, Livnat Y, Thraen I, Brody A, Rupper R. Evaluation of an electronic module for reconciling medications in home health plans of care. Appl Clin Inform 2017 Dec 16;07(02):412-424. [doi: 10.4338/aci-2015-11-ra-0154]

27. Wang H, Meng L, Song J, Yang J, Li J, Qiu F. Electronic medication reconciliation in hospitals: a systematic review and meta-analysis. Eur J Hosp Pharm 2018 Sep 08;25(5):245-250 [FREE Full text] [doi: 10.1136/ejhpharm-2017-001441] [Medline: 31157034]

28. Mekonnen AB, Abebe TB, McLachlan AJ, Brien JE. Impact of electronic medication reconciliation interventions on medication discrepancies at hospital transitions: a systematic review and meta-analysis. BMC Med Inform Decis Mak 2016 Aug 22;16(1):112 [FREE Full text] [doi: 10.1186/s12911-016-0353-9] [Medline: 27549581]

29. Marien S, Krug B, Spinewine A. Electronic tools to support medication reconciliation: a systematic review. J Am Med Inform Assoc 2017 Jan;24(1):227-240 [FREE Full text] [doi: 10.1093/jamia/ocw068] [Medline: 27301747]

30. Bassi J, Lau F, Bardal S. Use of information technology in medication reconciliation: a scoping review. Ann Pharmacother 2010 May 01;44(5):885-897. [doi: 10.1345/aph.1m699]

31. Kripalani S, Hart K, Schaninger C, Bracken S, Lindsell C, Boyington DR. Use of a tablet computer application to engage patients in updating their medication list. Am J Health Syst Pharm 2019 Feb 09;76(5):293-300 [FREE Full text] [doi: 10.1093/ajhp/zxy047] [Medline: 30753287]

32. Marien S, Legrand D, Ramdoyal R, Nsenga J, Ospina G, Ramon V, et al. A user-centered design and usability testing of a web-based medication reconciliation application integrated in an eHealth network. Int J Med Inform 2019 Jun;126:138-146. [doi: 10.1016/j.ijmedinf.2019.03.013] [Medline: 31029255]

33. Lesselroth B, Adams S, Felder R, Dorr DA, Cauthers P, Church V, et al. Using consumer-based kiosk technology to improve and standardize medication reconciliation in a specialty care setting. Jt Comm J Qual Patient Saf 2009 May;35(5):264-AP1. [doi: 10.1016/s1553-7250(09)35037-0]

34. Lesselroth B, Adams K, Church V, Tallett S, Russ Y, Wiedrick J, et al. Evaluation of multimedia medication reconciliation software: a randomized controlled, single-blind trial to measure diagnostic accuracy for discrepancy detection. Appl Clin Inform 2018 Apr 02;9(2):285-301 [FREE Full text] [doi: 10.1055/s-0038-1645889] [Medline: 29719884]

35. Wright T, Adams K, Church V, Ferraro M, Ragland S, Sayers A, et al. Implementation of a medication reconciliation assistive technology: a qualitative analysis. AMIA Annu Symp Proc 2017;2017:1802-1811 [FREE Full text] [Medline: 29854251]

XSL·FO

RenderX

Abbreviations

BPMH: best-possible medication history
cRCT: cluster-randomized controlled trial
EHR: electronic health record
IRR: incidence rate ratio
MT: MedTrue
OR: odds ratio
UC: usual care

XSL•FO
RenderX

Original Paper

Weight Loss Trajectories in Healthy Weight Coaching: Cohort Study

Sakris K E Kupila[1*], BM; Mikko S Venäläinen[2*], PhD; Laura-Unnukka Suojanen[3], MSc; Milla Rosengård-Bärlund[3], MD, PhD; Aila J Ahola[1,3,4,5], PhD; Laura L Elo[2], PhD; Kirsi H Pietiläinen[1,3], MD, PhD

[1]Obesity Research Unit, Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[2]Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland

[3]Abdominal Center, Obesity Center, Endocrinology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland

[4]Abdominal Center, Nephrology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland

[5]Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland

[*]these authors contributed equally

Corresponding Author:
Sakris K E Kupila, BM
Obesity Research Unit, Research Program for Clinical and Molecular Metabolism
Faculty of Medicine
University of Helsinki
P.O. Box 63
University of Helsinki
Helsinki, FI-00014
Finland
Phone: 358 02941911
Email: sakris.kupila@helsinki.fi

Abstract

Background: As global obesity prevalence continues to increase, there is a need for accessible and affordable weight management interventions, such as web-based programs.

Objective: This paper aims to assess the outcomes of healthy weight coaching (HWC), a web-based obesity management program integrated into standard Finnish clinical care.

Methods: HWC is an ongoing, structured digital 12-month program based on acceptance and commitment therapy. It includes weekly training sessions focused on lifestyle, general health, and psychological factors. Participants received remote one-on-one support from a personal coach. In this real-life, single-arm, prospective cohort study, we examined the total weight loss, weight loss profiles, and variables associated with weight loss success and program retention in 1189 adults (963 women) with a BMI >25 kg/m² among participants of the program between October 2016 and March 2019. Absolute (kg) and relative (%) weight loss from the baseline were the primary outcomes. We also examined the weight loss profiles, clustered based on the dynamic time-warping distance, and the possible variables associated with greater weight loss success and program retention. We compared different groups using the Mann-Whitney test or Kruskal-Wallis test for continuous variables and the chi-squared test for categorical variables. We analyzed changes in medication using the McNemar test.

Results: Among those having reached the 12-month time point (n=173), the mean weight loss was 4.6% (SE 0.5%), with 43% (n=75) achieving clinically relevant weight loss (≥5%). Baseline BMI ≥40 kg/m² was associated with a greater weight loss than a lower BMI (mean 6.6%, SE 0.9%, vs mean 3.2%, SE 0.6%; P=.02). In addition, more frequent weight reporting was associated with greater weight loss. No significant differences in weight loss were observed according to sex, age, baseline disease, or medication use. The total dropout rate was 29.1%. Dropouts were slightly younger than continuers (47.2, SE 0.6 years vs 49.2, SE 0.4 years; P=.01) and reported their weight less frequently (3.0, SE 0.1 entries per month vs 3.3, SE 0.1 entries per month; P<.001).

Conclusions: A comprehensive web-based program such as HWC is a potential addition to the repertoire of obesity management in a clinical setting. Heavier patients lost more weight, but weight loss success was otherwise independent of baseline characteristics.

Introduction

Global increases in the prevalence of obesity represent a pressing public health concern. Obesity is closely linked to various physical and mental health impairments, including type 2 diabetes, hypertension, nonalcoholic fatty liver disease, cancer, depression, and a decreased health-related quality of life, to name a few [1]. Among individuals with obesity, a moderate weight loss of 5% carries significant metabolic health benefits [2]. Furthermore, the etiology of obesity is multifactorial and consists of genetic, environmental, and behavioral factors. Behavioral factors, such as appetite control, eating habits, and physical activity, remain central targets in clinical practice and substantially impact treatment outcomes [3].

Treating obesity traditionally relies on lifestyle modifications during individual or group face-to-face sessions [4]. With ever-increasing rates of obesity and limited financial resources within health care systems, we must identify effective weight loss interventions that can be delivered to the wider public at a reasonable cost. Using information technology in obesity treatment may offer a cheaper alternative to relying on employed staff alone. Furthermore, while setting up an internet-based program may be expensive, upkeep can be cost-effective [5,6]. In addition, remote guidance may also reduce the number of necessary in-person consultations and minimize the travel time and costs associated with location-dependent treatment. Financial benefits and technology-based programs can be accessible regardless of the time of day, participants' life situations, places of residence, and potential disabilities or oral communication difficulties [7].

Common successful components of web-based interventions include self-monitoring, professional feedback, goal-setting, social support, and a structured program [5,8-10]. Tailoring programs in this manner may enhance self-efficacy, evoke a feeling of being cared for, improve end results, increase engagement, and decrease attrition [11-14]. Furthermore, web-based programs may enhance the feeling of control over one's own care and facilitate patient–expert and patient–patient interactions [11,15].

The effectiveness of web-based weight loss and weight management programs can diminish because of the poor use of the program and its elements. Reasons for this may include outdated website design, insufficient internet skills, limited internet connection, and patient unfamiliarity with web-based interventions in health care [15-17]. Moreover, users may perceive programs requiring self-monitoring to be burdensome [13]. Additional obstacles in technology-assisted programs may include problems with confidentiality and privacy, a decrease in face-to-face communication, medical errors caused by system malfunctions, technological errors, and data manipulation and misinterpretation, as well as legal, ethical, and administrative barriers [13,17]. Finally, high attrition rates and decreasing engagement over time represent common features of both eHealth innovations and weight loss interventions in general [18-20].

Although obesity treatment remains multidisciplinary, previously published studies typically focused on a limited number of factors important to successful weight loss. Most available studies on web-based programs offered no real-time human support to participants, and only a few studies have been conducted in real-life clinical settings in diverse patient groups. Moreover, previous studies have primarily examined small populations with short intervention and follow-up periods [21,22]. However, there are a few important exceptions. In a previous Danish study conducted in a real-life municipal setting, an eHealth intervention resulted in a significant weight reduction of 4.3% at the mean 7.3-month time point [23]. In patients who remained in the intervention for over 9 months, the mean weight loss was 6.3% [23]. Thus, although eHealth interventions are promising alternatives to conventional interventions, it remains important to assess their benefits in actual clinical care.

In this paper, we report our initial observations from a large cohort of patients taking part in a web-based platform, HealthyWeightHub, through which patients use an interactive program, Healthy Weight Coaching (HWC), for obesity management. The cohort consisted of men and women with a wide age and BMI range. Furthermore, the program relies on a broad spectrum of approaches (eg, diet, physical activity, sleep, health, psychological factors, and coping with stress) in the form of web-based training modules supported by a personal coach. In this real-life, single-arm, prospective cohort study, we assessed the outcomes of the HWC program. In doing so, we focused on the amount of weight loss, interindividual variability in the weight loss response, and possible variables associated with weight loss success and program retention.

Methods

Intervention

HWC is an ongoing 12-month interactive web-based intervention for weight management among patients identified with overweight or obesity. The program is available free of charge to all Finnish citizens as a part of Finnish public health care. In the program, we addressed a broad spectrum of health behaviors related to weight management, including diet, physical activity, sleep, psychological factors, coping with stress, and general health status. The program structure relied on weekly training sessions, and participants can freely choose from 200 available sessions to best meet their individual needs. Participants could submit daily—and were instructed to do so at least weekly—their weight, targets, feelings, diet, and physical activity logs to the program. In addition, each patient was assigned a personal coach (a nurse, nutritionist, physiotherapist, or psychologist) who offered remote one-on-one support and could individually tailor specific sessions to patients.

One-on-one coaching was arranged biweekly during the first month, after which it was arranged monthly. Alongside coaching, participants were able to exchange messages with the coach whenever needed. Participants could also interact with each other using anonymous group chats. The group chats were arranged monthly. To ensure cybersecurity, the program required strong authentication.

The HWC program and its training sessions are based on the framework of acceptance and commitment therapy (ACT) and other theories of behavior change. ACT is a form of cognitive behavioral therapy that supports flexible decision-making in everyday life. Specifically, ACT increases mindfulness, self-regulation, and psychological flexibility [24-27]. In addition, the elements incorporated include self-monitoring, counselor feedback and communication, group support, a structured program, and individual tailoring, all of which appear to support successful weight loss [14,28,29]. The program has been previously described in detail [30].

Participants

Patients entered the program based on a referral from a licensed physician in Finland. Most of the patients were from the Hospital District of Helsinki and Uusimaa. The general inclusion criteria in the referral process for the HWC were as follows: (1) age ≥18 years, (2) BMI ≥25 kg/m², (3) access to a computer or a smartphone with a stable internet connection, and (4) a willingness and motivation to participate in a web-based treatment program. Persons who were pregnant, lactating, or for other reasons required face-to-face treatment modalities were not accepted to the program, as their intervention would be conducted more safely in the clinics rather than via the internet.

Ethical Considerations

Each participant provided written informed consent. The study was approved by the Coordinating Ethics Committee of the Helsinki University Hospital (reference number 327/13/03/00/2015).

Data Collection

The database for this study was created on March 12, 2019, and included data from all consenting participants since the initiation of the HWC in October 2016 [30]. To avoid confounding factors, we excluded participants who were diagnosed with type 1 diabetes; were following a very low-calorie diet; were taking weight loss medication; had undergone a gastric balloon procedure or bariatric surgery; and those who did not complete a 2-week trial period (n=264, 18.2%). All other patients who were initially referred to the HWC were included in the analyses.

We obtained data regarding age and sex from the Finnish national register. During the program, the participants completed several internet-based questionnaires, as previously described in detail [30]. To summarize, height and weight as well as baseline morbidities and medications were reported upon entry into the program. Daily reporting for weight was optional. However, at a minimum, weekly reports were encouraged as part of the training sessions. We used these weekly reports to interpolate the daily weight measurements. We calculated BMI

as weight in kilograms divided by height in meters squared (kg/m²) and dichotomized it into groups with BMI <40 kg/m² (n=557) and BMI ≥40 kg/m² (n=631). This grouping at the morbid obesity cut-off point was used instead of conventional overweight (BMI 25-29.9 kg/m²) and obesity (BMI ≥30 kg/m²) groups, as only 25 individuals were identified in the overweight group. We also calculated the entry rate (entries per month) from the total number of weight entries divided by the time until the last weight entry. We considered participants who were inactive for >90 days as dropouts from the program.

Statistical Analyses

As recruitment to the program remains continuous, the number of participants reaching each time point varied (Multimedia Appendix 1, Table S1). Unless otherwise stated, the figures and statistical analyses were based on the data from all participants. Data are presented as frequencies (%) for categorical variables and mean (SE) and median (IQR) for continuous variables. All continuous variables had skewed distributions (Shapiro-Wilk normality test). All statistical analyses were performed using the R statistical computing environment (version 3.4.1; R Foundation for Statistical Computing) [31]. Group comparisons were conducted using the Mann-Whitney test (2 groups) or Kruskal-Wallis test (>2 groups) for continuous variables and the chi-squared test for categorical variables. Changes in medication between baseline and 12 months were analyzed using the McNemar test. Weight loss patterns were clustered based on the dynamic time-warping distance and agglomerative hierarchical clustering using the R package *dtwclust* [32]. The R package *ggplot2* was used for the visualization of all results [33]. We set the level of significance at $P<.05$.

Results

Baseline Characteristics

Table 1 summarizes the baseline characteristics of the 1189 participants in this study. The mean age of patients was 48.6 (SE 0.3) years. BMI ranged from 26.3 to 78.7 kg/m² (mean 40.6, SE 0.2 kg/m²). There were no statistically significant differences in age or BMI between men and women.

In total, 1044 (87.8%) participants reported comorbidities of some type. The mean number of comorbidities reported per person was 4.8 (SE 0.1; range 0-15). Hypertension was the most prevalent comorbidity (n=598, 57.2%), followed by dyslipidemia (n=368, 35.2%), allergies (n=368, 35.2%), sleep apnea (n=358, 34.2%), depression (n=343, 32.8%), and osteoarthritis or osteoarthrosis (n=318, 30.4%; Multimedia Appendix 1, Table S2). In addition, 18.5% (n=193) of the participants reported type 2 diabetes as a comorbidity.

Altogether, 1000 (84.1%) participants reported using any medication. The reported medications were grouped as shown in Multimedia Appendix 1, Table S3. The mean number of these medication groups per patient was 2.5 (SE 0.1, range 0-10). The most prevalent medication groups were cardiovascular drugs (n=513, 43.1%), followed by dietary supplements (n=434, 36.5%), metabolic and endocrine drugs (n=349, 29.4%), acute pain medication (n=348, 29.3%), and psychopharmacological drugs (n=284, 23.9%; Multimedia Appendix 1, Table S3). At

12 months (n=173), the mean number of medication groups remained the same (2.5, SE 0.03), whereas the use of respiratory and acute pain medication changed significantly (Multimedia Appendix 1, Table S4).

Table 1. Basic characteristics of the study population.

Characteristic	All (N=1189)		Women (n=963)		Men (n=226)	
	Value, mean (SE; range)	Value, median (IQR)	Value, mean (SE; range)	Value, median (IQR)	Value, mean (SE; range)	Value, median (IQR)
Age (years)	48.6 (0.3; 19.0-78.0)	50.0 (17.0)	48.4 (0.4; 19.0-78.0)	50.0 (17.0)	49.6 (0.8; 22.0-73.0)	52.0 (19.0)
Height (cm)	168.7 (0.2; 147.0-198.0)	168.0 (10.0)	166.0 (0.2; 147.0-184.0)	166.0 (8.0)	180.3 (0.5; 158.0-198.0)	180.0 (10.0)
Weight (kg)	115.8 (0.7; 60.2-284.0)	112.5 (30.2)	111.3 (0.7; 60.2-193.0)	108.0 (26.2)	135.2 (1.8; 83.0-284.0)	130.0 (32.2)
BMI (kg/m^2)	40.6 (0.2; 26.3-78.7)	39.6 (8.3)	40.4 (0.2; 26.3-72.5)	39.4 (8.2)	41.5 (0.5; 27.2-78.7)	40.2 (8.6)

Mean and Categorical Weight Loss

The mean weight loss trajectory throughout the 12-month intervention period is shown in Figure 1A. At 12 months, the mean weight loss reached 4.6% (SE 0.5%) or 5.6 kg (SE 0.7 kg). The mean relative (%) and absolute (kg) weight losses at 3, 6, 9, and 12 months are shown in Multimedia Appendix 1, Table S5. Figure 1B shows the categorical weight loss at each time point. Among all participants who reached the 12-month time point by the data lock (n=173), 12.1% (n=21) lost 3% to 4.9%, 27.7% (n=48) lost 5% to 9.9%, and 15.6% (n=27) lost ≥10% of their baseline body weight. Altogether, 43.3% (n=75) of participants reaching the 12-month time point reported a clinically relevant weight loss of ≥5% of their initial weight.

At 12 months (n=173), the mean weight loss was similar in men (mean 5.4%, SE 1.4%) and women (mean 4.5%, SE 0.6%; Multimedia Appendix 1, Table S6). Participants aged <40 years achieved a weight loss (mean 4.3%, SE 1.2%) comparable with that of older individuals (mean 4.7%, SE 0.6%). Participants with a baseline BMI ≥40 kg/m² exhibited greater weight loss than those with a baseline BMI <40 kg/m² (mean 6.6%, SE 0.9% vs mean 3.2%, SE 0.5%; P=.02). Instead, patients with and without type 2 diabetes achieved similar weight loss rates (mean 5.3%, SE 1.2% vs mean 4.2%, SE 0.6%; P=.45, respectively). By the end of the intervention, patients who reported their body weight ≥4 times per month did not achieve statistically significantly greater weight loss (mean 5.9%, SE 1.1%) than those who reported a body weight <4 times per month (mean 4.2%, SE 0.6%; P=.09). However, at 3-, 6-, and 9-month time points, significantly greater weight loss was observed for patients who reported their body weights more frequently (P<.001).

Figure 1. (A) Mean weight change (percentage from baseline) through the intervention; (B) distribution of weight loss categories at 3 (n=839), 6 (n=563), 9 (n=359), and 12 (n=173) months. The numbers indicate participants who have reached each time point.
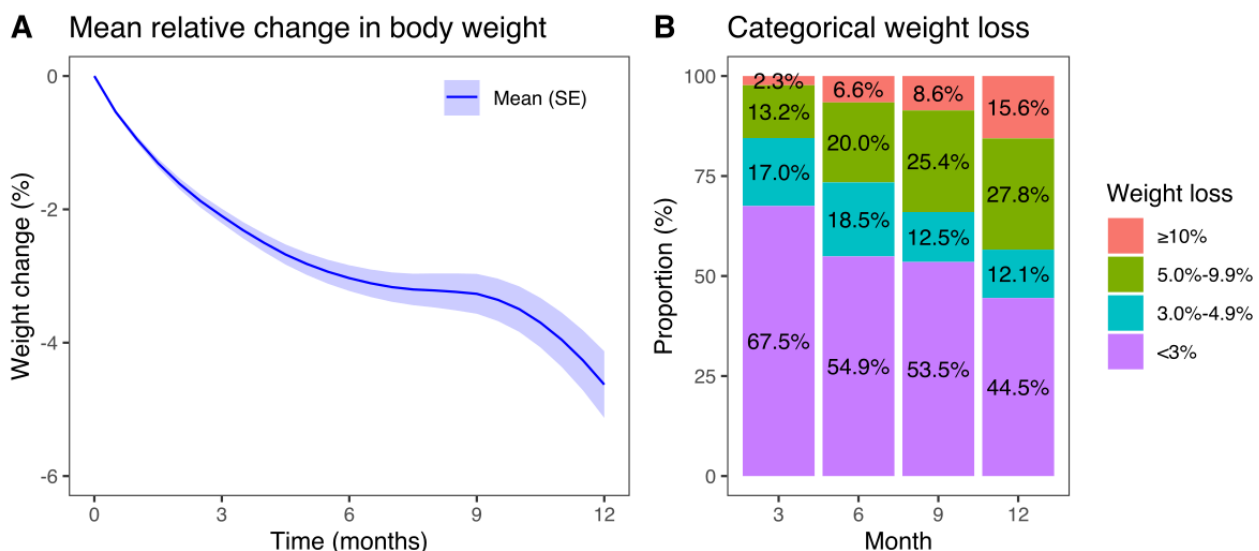
Interindividual Variation in Weight Loss Responses

In our attempt to delineate the interindividual variation in weight loss responses, we first plotted the last available observation from each participant, observing a wide range of weight changes, from −34.3% to +14.4% (Figure 2). Similarly, wide variations were observed at 3, 6, 9, and 12 months (Multimedia Appendix 1, Figure S1).

We next analyzed the variation in weight changes using a dynamic time-warping algorithm and hierarchical clustering. In this process, we identified 5 clusters into which participants could be divided based on their weight loss success (Figure 3).

The mean weight changes in each of the 5 weight loss clusters from baseline to 3, 6, 9, and 12 months are shown in Multimedia Appendix 1, Table S7. Those in cluster 1 (superresponders; n=93, 8% of the study population) lost weight rapidly and consistently until the end of the 12-month period. In this group, participants lost an average of 15.7% (SE 1.3%) of their baseline body weight at 12 months.

In cluster 2 (responders; n=208, 17.5% of the study population), participants lost weight rapidly during the first 3 months, after which the weight loss rate slowed down. In this group, the participants had lost an average of 6.1% (SE 0.3%) of their baseline body weight at 12 months.

In cluster 3 (moderate responders; n=332, 27.9% of the study population), participants exhibited a small and steady mean 12-month weight loss of 3.4% (SE 0.3%) of their baseline body weight.

In cluster 4 (nonresponders; n=384, 32.3% of the study population), neither significantly lost nor gained weight. In this group, participants had lost on average 0.1% (SE 0.2%) of their baseline body weight at 12 months.

In cluster 5 (gainers; n=172, 14.5% of the study population), participants slowly gained weight rather than losing it. In this group, participants gained on average 3.5% (SE 0.6%) of their baseline body weight at 12 months.

Across all participants, 53.2% (n=633) fell into the category of responders (clusters 1-3), 32.3% (n=384) represented nonresponders (cluster 4), and 14.5% (n=172) represented gainers (cluster 5).

Figure 2. Individual weight change (percentage from baseline) at the last available time point.
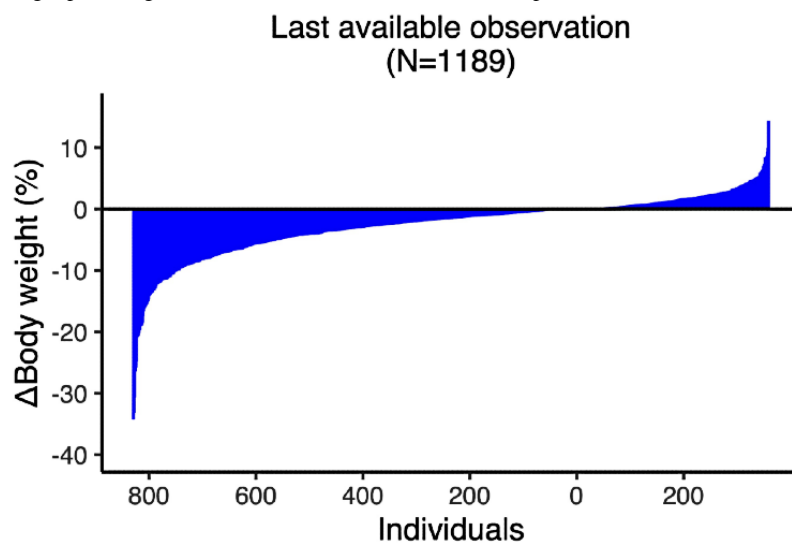
Figure 3. Clusters of study participants (N=1189) based on weight loss success trajectories. Individual weight loss patterns were clustered based on dynamic time-warping distance and agglomerative hierarchical clustering. Each colored line represents the weight change trajectory of a participant. The black line represents the average weight change in said cluster. Cluster 1, superresponders; cluster 2, responders; cluster 3, moderate responders; cluster 4, nonresponders; cluster 5, gainers.
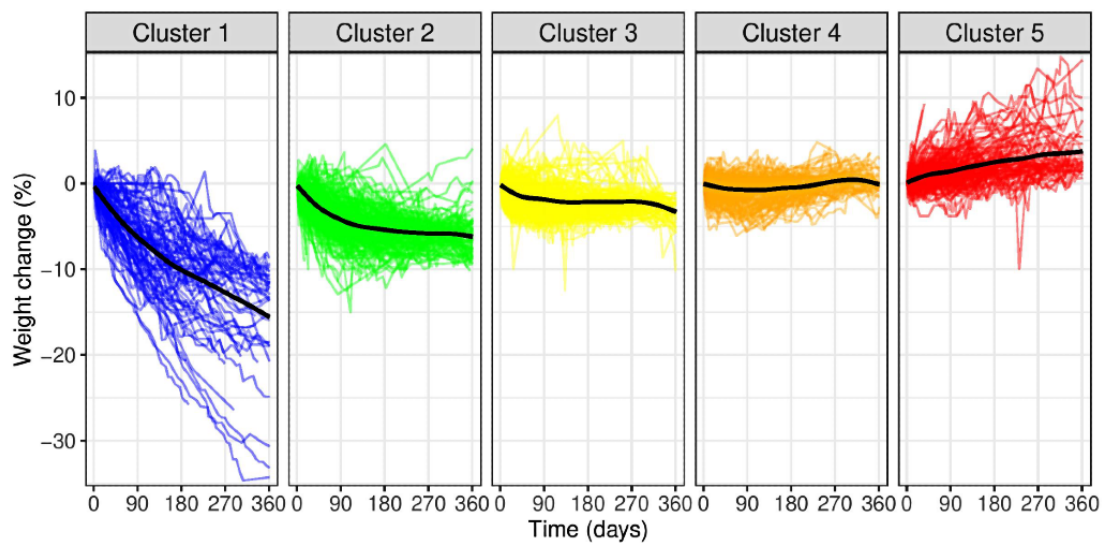
Factors Explaining the Interindividual Variation in Weight Loss Response

To identify factors contributing to weight loss success, we analyzed whether the clusters differed according to age, sex, baseline BMI, concomitant diseases or medication use, or the frequency of weight entries. The clusters exhibited no statistically significant differences in age or sex (Multimedia Appendix 1, Table S8). Instead, a higher baseline BMI was associated with greater mean weight loss success ($P<.001$). Furthermore, greater mean weight loss was associated with more frequent weight entries ($P=.004$). In addition, the clusters differed only in relation to the number of medications ($P=.048$) but not in the medication groups and diseases (Multimedia Appendix 1, Table S9 and Figure S2).

Attrition

Finally, we analyzed the attrition rate and possible associated factors. The total dropout rate at 12 months was 29.1% (n=346; Multimedia Appendix 1, Figure S3). In addition, given the continuous recruitment, 56.6% (n=689) of participants had not yet reached the 12-month time point at the time of the data lock.

We found no significant differences in the baseline BMI (mean 41.0 kg/m^2, SE 0. 4 kg/m^2 vs mean 40.4 kg/m^2, SE 0.2 kg/m^2; $P=.12$) or sex (272, 79% vs 691, 82% women; $P=.21$) between those who prematurely discontinued the program and those who remained continuously active (Multimedia Appendix 1, Table S10). Those discontinuing were younger (mean 47.2 years, SE 0.6 years vs 49.2 years, SE 0.4 years; $P=.01$) and made fewer monthly weight entries (mean 3.0, SE 0.1 entries/month vs mean 3.3, SE 0.1 entries per month; $P<.001$) compared with those who adhered. On the basis of their last reported body weight, most individuals who dropped out, clustered either as nonresponders or gainers (Multimedia Appendix 1, Figure S4).

Discussion

Principal Findings

The data collected in the novel web-based weight management program, HWC, offered a unique perspective of real-time obesity management in a large population of individuals with a wide age and BMI range. They also enabled us to identify subgroups of patients with different weight trajectories and trends in weight loss success. Furthermore, with a large number of included modules (including diet, physical activity, sleep, psychological factors, stress, and general health), the HWC program encompasses the multidisciplinary nature of clinical weight management with a long intervention time. Importantly, our patients, on average, presented with morbid obesity, with several comorbidities and the use of various medications. HWC is embedded into standard Finnish clinical care, and thus represents real-life clinical patient material. However, although the data collection was preplanned alongside the routine delivery of care, one needs to bear in mind that this real-life study has no control group.

We observed a mean weight loss of 4.6% (SE 0.5%) or 5.6 kg (SE 0.7 kg) at the 12-month time point. In a meta-analysis of 15 personalized eHealth studies, Lau et al [21] reported a 2.8

kg (range 2.0 kg-3.5 kg) greater mean weight loss in the eHealth intervention group than in the control group. Similarly, in a systematic review, a 2.4 kg higher weight loss was observed in traditional face-to-face treatments compared with no or minimal interventions [34]. In a previous Finnish randomized controlled trial, patients treated only via a web-based behavioral weight loss program lost on average 1.2% (range 0.3% to 2.2%) of their baseline body weight in 12 months, whereas the group treated via both the program and cognitive behavioral group therapy lost on average 3.5% (range 2.1% to 4.8%) of their baseline body weight [35]. In this real-life study, we did not have a control group, whereby we were unable to assess how a similar patient population to our study would have fared longitudinally without any treatment. On the basis of previous longitudinal studies, the BMI in Finnish cohorts tends to either increase [36] or remain stable [37] over time. Furthermore, in a Swedish Obese Subjects trial, where participants had a similar mean BMI (40.1 kg/m^2) as in our study, there was no weight change in the control group with usual care at the 12-month time point [38]. Thus, it seems likely that the mean 4.6% (SE 0.5%) weight loss we observed was a result of the HWC treatment. However, based on a previous randomized controlled trial [35], it is also possible that combining web-based treatment with face-to-face behavioral therapy would have been more effective than web-based treatment alone.

We identified 5 clusters based on weight loss success, when examining weight change trajectories, allowing us to explore characteristics predicting weight loss outcomes. The clusters were comparable in terms of age and sex. However, a higher baseline BMI was associated with greater weight loss success. We found no statistically significant differences in the number of comorbidities or medication groups among clusters, suggesting that medication use and concomitant diseases did not define features of weight loss success. Of note, cluster 1, which had the best weight loss success, also had the largest variability in the weight loss results, which is most likely because this cluster size was the smallest population. Moreover, this cluster had the highest baseline BMI values. The largest of the formed clusters, cluster 4, comprised individuals with no change in their body weight. Indeed, it is very typical that not all patients undergoing weight loss interventions are successful in losing weight [23,39]. Identifying individuals and the reasons for not responding would be important to better support these individuals in future interventions. Overall, 53.2% (n=633) of the participants fell into clusters 1 to 3, and thus experienced some weight loss. Moreover, the observed mean weight loss was 4.6% (SE 0.5%), which is comparable with or somewhat higher than that previously reported [21,34,35].

No difference in weight loss was observed between individuals with and those without type 2 diabetes. Interestingly, in previous weight loss studies, patients with diabetes exhibited poorer success than those without diabetes [40,41]. One reason might be that diabetes as a comorbidity can render weight loss more difficult and harder to maintain [40,42]. However, diabetes and its related comorbidities may result in greater adherence and motivation for weight loss [43]. Furthermore, these novel interactive approaches may be more engaging and empowering than traditional methods because web-based programs may

enhance a feeling of self-control and accountability regarding one's care [15].

We found that a higher number of weight entries was associated with greater weight loss, in agreement with previous findings [10,44,45]. We also found that those who dropped out made fewer weight entries to the program than did those who adhered. However, whether the number of entries attributes to success itself remains unclear. Success may make the program more rewarding, thus prompting a patient to log weights more frequently and engage more with the program. As for attrition, it is probable that individuals who fail to lose weight or experience any other benefits more readily drop out from the programs. In addition, discontinuing an intervention may also occur because the participants already met their personal goals before the targeted end date [17,46-48]. Understanding the nature of this relationship between success and the weight entry rate requires further investigation.

In the HWC program, we observed an attrition rate of 29.1% (n=346). Patients who dropped out were younger and more often belonged to the group that did not lose weight in the program. Retention and weight loss may be bidirectionally related to each other: patients retained in the program have more support for weight loss attempts, and better weight loss motivates them to remain active in the program. Overall, attrition may bias the outcome and should be considered when interpreting the results of the study. High attrition rates and decreasing engagement over time represent common features of both eHealth innovations and clinical face-to-face weight loss interventions [18-20]. This may be especially true in long-term interventions, such as this program. Similar to our study, the 2 eHealth platforms for weight management reported a 54% attrition rate in a clinical setting [35,46]. This is comparable with a traditional face-to-face clinic-based intervention, which also reported that 54% of patients discontinued [19]. Previous meta-analyses on eHealth-based weight loss interventions primarily focused on randomized controlled trials and determined average attrition rates close to 20% [6,10]. However, these rates might misrepresent attrition in real-life programs given inherent differences in, for example, participant demographics (ie, real clinical patients vs volunteers) and program design (ie, additional assessment sessions, participant retention strategies, and greater accountability).

This study had several strengths and limitations. An important strength of this study lay in its reliance on a large sample of individuals with a broad age and BMI distribution. With a large number of reported comorbidities and medications, our data provided a realistic view of treating patients with obesity-related complications. Although using data from a real-life clinical setting offers a tangible and practical view of the web-based management of obesity, the program was not randomized and was thus lacking a control group. Subsequently, the study has low internal validity, and we are unable to determine the efficacy of the program compared with no treatment, face-to-face treatment, or any other comparison. We were also not able to study those patients who needed treatment for obesity but were not willing to participate in a web-based program and received a referral to HWC in the first place. In addition, as body weight was self-reported, the possibility of misreporting will need to be considered when interpreting the results. It should also be noted that the program comprises several components. On the basis of the current analyses, it is not possible to distinguish between the different treatment effects, dose responses, or mechanisms of change in these components. The current report describes the overall outcomes of this program. Finally, as is the case with most previous studies, our participants consisted predominantly of women and adults of middle age. Therefore, our conclusions are not fully generalizable to younger adults and men.

Future research should focus on younger adults and men, in particular, and attempt to examine the effects and appeal of web-based programs. In addition, further investigation of the factors associated with successful weight loss and reasons for attrition would prove beneficial. For example, future studies should attempt to determine when attrition occurs because of reaching weight loss targets early instead of genuine nonuse. This would advance the development of targeted web-based programs tailored to specific patient subgroups.

Conclusions

In conclusion, HWC is a web-based method for the management of obesity in real-life clinical settings. During the 12-month treatment, the average weight loss was 4.6% (SE 0.5%) but varied widely. Specifically, baseline medication use, baseline health status, age, or sex was not significantly associated with weight loss in the HWC program. Further research is needed to understand the determinants of weight loss success.

XSL•FO

RenderX

Authors' Contributions

SKEK was a major contributor in writing the manuscript. MSV analyzed and interpreted the data, created figures and tables, and revised the manuscript. LUS was a major contributor to the conception, study design, and data acquisition. MRB contributed to data acquisition and refinement. AJA substantively revised the manuscript. LLE contributed to data analysis and presentation. KHP was a major contributor to the conception, study design, data acquisition, and supervision of the study. All authors reviewed and approved the manuscript before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables and figures.
[PDF File (Adobe PDF File), 626 KB - formative_v6i3e26374_app1.pdf]

References

1. Bray GA. Medical consequences of obesity. J Clin Endocrinol Metab 2004 Jun;89(6):2583-2589. [doi: 10.1210/jc.2004-0535] [Medline: 15181027]
2. Magkos F, Fraterrigo G, Yoshino J, Luecking C, Kirbach K, Kelly S, et al. Effects of moderate and subsequent progressive weight loss on metabolic function and adipose tissue biology in humans with obesity. Cell Metab 2016 Apr 12;23(4):591-601 [FREE Full text] [doi: 10.1016/j.cmet.2016.02.005] [Medline: 26916363]
3. Castelnuovo G, Pietrabissa G, Manzoni GM, Corti S, Ceccarini M, Borrello M, et al. Chronic care management of globesity: promoting healthier lifestyles in traditional and mHealth based settings. Front Psychol 2015;6:1557 [FREE Full text] [doi: 10.3389/fpsyg.2015.01557] [Medline: 26528215]
4. Wadden TA, Butryn ML, Wilson C. Lifestyle modification for the management of obesity. Gastroenterology 2007 May;132(6):2226-2238. [doi: 10.1053/j.gastro.2007.03.051] [Medline: 17498514]
5. Manzoni GM, Pagnini F, Corti S, Molinari E, Castelnuovo G. Internet-based behavioral interventions for obesity: an updated systematic review. Clin Pract Epidemiol Ment Health 2011 Mar 04;7:19-28 [FREE Full text] [doi: 10.2174/1745017901107010019] [Medline: 21552423]
6. Kodama S, Saito K, Tanaka S, Horikawa C, Fujiwara K, Hirasawa R, et al. Effect of Web-based lifestyle modification on weight control: a meta-analysis. Int J Obes (Lond) 2012 May;36(5):675-685. [doi: 10.1038/ijo.2011.121] [Medline: 21694698]
7. Kersting A, Schlicht S, Kroker K. [Internet therapy. Opportunities and boundaries]. Nervenarzt 2009 Jul;80(7):797-804. [doi: 10.1007/s00115-009-2721-5] [Medline: 19404602]
8. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015 May;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]
9. Semper HM, Povey R, Clark-Carter D. A systematic review of the effectiveness of smartphone applications that encourage dietary self-regulatory strategies for weight loss in overweight and obese adults. Obes Rev 2016 Sep;17(9):895-906. [doi: 10.1111/obr.12428] [Medline: 27192162]
10. Neve M, Morgan P, Jones P, Collins C. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. Obes Rev 2010 Apr;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x] [Medline: 19754633]
11. Coulter A, Entwistle V, Eccles A, Ryan S, Shepperd S, Perera R. Personalised care planning for adults with chronic or long-term health conditions. Cochrane Database Syst Rev 2015 Mar 03(3):CD010523 [FREE Full text] [doi: 10.1002/14651858.CD010523.pub2] [Medline: 25733495]
12. Lyzwinski LN, Caffery LJ, Bambling M, Edirippulige S. Consumer perspectives on mHealth for weight loss: a review of qualitative studies. J Telemed Telecare 2017 Feb 09;24(4):290-302. [doi: 10.1177/1357633x17692722]
13. Aalbers T, Baars M, Rikkert MO. Characteristics of effective internet-mediated interventions to change lifestyle in people aged 50 and older: a systematic review. Ageing Res Rev 2011 Sep;10(4):487-497. [doi: 10.1016/j.arr.2011.05.001] [Medline: 21628005]
14. Ryan K, Dockray S, Linehan C. A systematic review of tailored eHealth interventions for weight loss. Digit Health 2019;5:2055207619826685 [FREE Full text] [doi: 10.1177/2055207619826685] [Medline: 30783535]
15. Sorgente A, Pietrabissa G, Manzoni GM, Re F, Simpson S, Perona S, et al. Web-based interventions for weight loss or weight loss maintenance in overweight and obese people: a systematic review of systematic reviews. J Med Internet Res 2017 Jun 26;19(6):e229 [FREE Full text] [doi: 10.2196/jmir.6972] [Medline: 28652225]
16. Levine DM, Savarimuthu S, Squires A, Nicholson J, Jay M. Technology-assisted weight loss interventions in primary care: a systematic review. J Gen Intern Med 2015 Jan;30(1):107-117 [FREE Full text] [doi: 10.1007/s11606-014-2987-6] [Medline: 25134692]

XSL•FO
RenderX

17. Lie SS, Karlsen B, Oord ER, Graue M, Oftedal B. Dropout from an eHealth intervention for adults with type 2 diabetes: a qualitative study. J Med Internet Res 2017 May 30;19(5):e187 [FREE Full text] [doi: 10.2196/jmir.7479] [Medline: 28559223]

18. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

19. Gill R, Karmali S, Hadi G, Al-Adra D, Shi X, Birch D. Predictors of attrition in a multidisciplinary adult weight management clinic. Can J Surg 2012 Aug;55(4):239-243 [FREE Full text] [doi: 10.1503/cjs.035710] [Medline: 22617538]

20. Moroshko I, Brennan L, O'Brien P. Predictors of dropout in weight loss interventions: a systematic review of the literature. Obes Rev 2011 Nov;12(11):912-934. [doi: 10.1111/j.1467-789X.2011.00915.x] [Medline: 21815990]

21. Lau Y, Chee DG, Chow XP, Cheng LJ, Wong SN. Personalised eHealth interventions in adults with overweight and obesity: a systematic review and meta-analysis of randomised controlled trials. Prev Med 2020 Mar;132:106001. [doi: 10.1016/j.ypmed.2020.106001] [Medline: 31991155]

22. Rumbo-Rodríguez L, Sánchez-SanSegundo M, Ruiz-Robledillo N, Albaladejo-Blázquez N, Ferrer-Cascales R, Zaragoza-Martí A. Use of technology-based interventions in the treatment of patients with overweight and obesity: a systematic review. Nutrients 2020 Nov 26;12(12):3634 [FREE Full text] [doi: 10.3390/nu12123634] [Medline: 33255982]

23. Komkova A, Brandt CJ, Hansen Pedersen D, Emneus M, Sortsø C. Electronic health lifestyle coaching among diabetes patients in a real-life municipality setting: observational study. JMIR Diabetes 2019 Mar 12;4(1):e12140 [FREE Full text] [doi: 10.2196/12140] [Medline: 30860486]

24. Warren JM, Smith N, Ashwell M. A structured literature review on the role of mindfulness, mindful eating and intuitive eating in changing eating behaviours: effectiveness and associated potential mechanisms. Nutr Res Rev 2017 Jul 18;30(2):272-283. [doi: 10.1017/s0954422417000154]

25. Lillis J, Kendra KE. Acceptance and Commitment Therapy for weight control: model, evidence, and future directions. J Contextual Behav Sci 2014 Jan;3(1):1-7 [FREE Full text] [doi: 10.1016/j.jcbs.2013.11.005] [Medline: 25419510]

26. Sairanen E, Tolvanen A, Karhunen L, Kolehmainen M, Järvelä-Reijonen E, Lindroos S, et al. Psychological flexibility mediates change in intuitive eating regulation in acceptance and commitment therapy interventions. Public Health Nutr 2017 Apr 17;20(9):1681-1691. [doi: 10.1017/s1368980017000441]

27. Forman EM, Butryn ML, Manasse SM, Crosby RD, Goldstein SP, Wyckoff EP, et al. Acceptance-based versus standard behavioral treatment for obesity: results from the mind your health randomized controlled trial. Obesity (Silver Spring) 2016 Oct;24(10):2050-2056 [FREE Full text] [doi: 10.1002/oby.21601] [Medline: 27670400]

28. Willmott TJ, Pang B, Rundle-Thiele S, Badejo A. Weight management in young adults: systematic review of electronic health intervention components and outcomes. J Med Internet Res 2019 Feb 06;21(2):e10265 [FREE Full text] [doi: 10.2196/10265] [Medline: 30724736]

29. Raaijmakers LC, Pouwels S, Berghuis KA, Nienhuijs SW. Technology-based interventions in the treatment of overweight and obesity: a systematic review. Appetite 2015 Dec;95:138-151. [doi: 10.1016/j.appet.2015.07.008] [Medline: 26165415]

30. Suojanen L, Ahola A, Kupila S, Korpela R, Pietiläinen KH. Effectiveness of a web-based real-life weight management program: study design, methods, and participants' baseline characteristics. Contemp Clin Trials Commun 2020 Sep;19:100638 [FREE Full text] [doi: 10.1016/j.conctc.2020.100638] [Medline: 32885090]

31. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Core Team; 2013.

32. dtwclust: time series clustering along with optimizations for the dynamic time warping distance. CRAN. URL: https://CRAN.R-project.org/package=dtwclust [accessed 2022-02-05]

33. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Cham: Springer; 2016.

34. LeBlanc E, Patnode C, Webber E, Redmond N, Rushkin M, O'Connor E. Behavioral and pharmacotherapy weight loss interventions to prevent obesity-related morbidity and mortality in adults: updated evidence report and systematic review for the US preventive services task force. JAMA 2018 Sep 18;320(11):1172-1191. [doi: 10.1001/jama.2018.7777] [Medline: 30326501]

35. Teeriniemi A, Salonurmi T, Jokelainen T, Vähänikkilä H, Alahäivälä T, Karppinen P, et al. A randomized clinical trial of the effectiveness of a web-based health behaviour change support system and group lifestyle counselling on body weight loss in overweight and obese subjects: 2-year outcomes. J Intern Med 2018 Nov;284(5):534-545 [FREE Full text] [doi: 10.1111/joim.12802] [Medline: 29974563]

36. Roos E, Laaksonen M, Rahkonen O, Lahelma E, Lallukka T. Weight change and sickness absence--a prospective study among middle-aged employees. Eur J Public Health 2015 Apr;25(2):263-267. [doi: 10.1093/eurpub/cku087] [Medline: 24997201]

37. Konttinen H, van Strien T, Männistö S, Jousilahti P, Haukkala A. Depression, emotional eating and long-term weight changes: a population-based prospective study. Int J Behav Nutr Phys Act 2019 Mar 20;16(1):28 [FREE Full text] [doi: 10.1186/s12966-019-0791-8] [Medline: 30894189]

38. Sjöström L. Review of the key results from the Swedish Obese Subjects (SOS) trial - a prospective controlled intervention study of bariatric surgery. J Intern Med 2013 Mar;273(3):219-234 [FREE Full text] [doi: 10.1111/joim.12012] [Medline: 23163728]

XSL•FO
RenderX

39. Unick JL, Pellegrini CA, Demos KE, Dorfman L. Initial weight loss response as an indicator for providing early rescue efforts to improve long-term treatment outcomes. Curr Diab Rep 2017 Sep;17(9):69 [FREE Full text] [doi: 10.1007/s11892-017-0904-1] [Medline: 28726155]

40. Baker ST, Jerums G, Prendergast LA, Panagiotopoulos S, Strauss BJ, Proietto J. Less fat reduction per unit weight loss in type 2 diabetic compared with nondiabetic obese individuals completing a very-low-calorie diet program. Metabolism 2012 Jun;61(6):873-882. [doi: 10.1016/j.metabol.2011.10.017] [Medline: 22146094]

41. Antoun G, Nikpay M, McPherson R, Harper M, Dent R. Is type 2 diabetes in adults associated with impaired capacity for weight loss? Can J Diabetes 2018 Jun;42(3):313-6.e1. [doi: 10.1016/j.jcjd.2017.06.010] [Medline: 28800932]

42. Guare J, Wing R, Grant A. Comparison of obese NIDDM and nondiabetic women: short- and long-term weight loss. Obes Res 1995 Jul;3(4):329-335 [FREE Full text] [doi: 10.1002/j.1550-8528.1995.tb00158.x] [Medline: 8521149]

43. Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mHealth interventions for obesity and diabetes treatment and self-management. Adv Nutr 2017 May;8(3):449-462 [FREE Full text] [doi: 10.3945/an.116.014100] [Medline: 28507010]

44. Hunter CM, Peterson AL, Alvarez LM, Poston WC, Brundige AR, Haddock CK, et al. Weight management using the internet a randomized controlled trial. Am J Prev Med 2008 Feb;34(2):119-126. [doi: 10.1016/j.amepre.2007.09.026] [Medline: 18201641]

45. Harvey J, Krukowski R, Priest J, West D. Log often, lose more: electronic dietary self-monitoring for weight loss. Obesity (Silver Spring) 2019 Mar;27(3):380-384 [FREE Full text] [doi: 10.1002/oby.22382] [Medline: 30801989]

46. Pedersen DH, Mansourvar M, Sortsø C, Schmidt T. Predicting dropouts from an electronic health platform for lifestyle interventions: analysis of methods and predictors. J Med Internet Res 2019 Sep 04;21(9):e13617 [FREE Full text] [doi: 10.2196/13617] [Medline: 31486409]

47. Neve MJ, Collins CE, Morgan PJ. Dropout, nonusage attrition, and pretreatment predictors of nonusage attrition in a commercial web-based weight loss program. J Med Internet Res 2010 Dec 14;12(4):e69 [FREE Full text] [doi: 10.2196/jmir.1640] [Medline: 21156470]

48. Van der Mispel C, Poppe L, Crombez G, Verloigne M, De Bourdeaudhuij I. A self-regulation-based eHealth intervention to promote a healthy lifestyle: investigating user and website characteristics related to attrition. J Med Internet Res 2017 Jul 11;19(7):e241 [FREE Full text] [doi: 10.2196/jmir.7277] [Medline: 28698168]

Abbreviations

ACT: acceptance and commitment therapy
HWC: Healthy Weight Coaching

<u>Original Paper</u>

Benefits of mHealth Co-design for African American and Hispanic Adults: Multi-Method Participatory Research for a Health Information App

Devlon N Jackson[1,2*], MPH, PhD; Neil Sehgal[3*], MPH, PhD; Cynthia Baur[1,2*], PhD

[1]Department of Behavioral and Community Health, Center for Health Literacy, Center for Health Equity, School of Public Health, University of Maryland, College Park, MD, United States

[2]Department of Behavioral and Community Health, Center for Health Literacy, School of Public Health, University of Maryland, College Park, MD, United States

[3]Department of Health and Policy Management, School of Public Health, University of Maryland, College Park, MD, United States

[*]all authors contributed equally

Corresponding Author:
Devlon N Jackson, MPH, PhD
Department of Behavioral and Community Health, Center for Health Literacy, Center for Health Equity
School of Public Health
University of Maryland
4200 Valley Drive
College Park, MD, 20742
United States
Phone: 1 3014051464
Email: djacks04@umd.edu

Abstract

Background: Participatory research methodologies can provide insight into the use of mobile health (mHealth) apps, cultural preferences and needs, and health literacy issues for racial and ethnic groups, such as African Americans and Hispanics who experience health disparities.

Objective: This methodological paper aims to describe a 1-year multi-method participatory research process that directly engaged English-speaking African American and bilingual or Spanish-speaking Hispanic adults in designing a prevention-focused, personalized mHealth, information-seeking smartphone app. We report design team participants' experiences with the methods to show why our approach is valuable in producing apps that are more aligned with their needs.

Methods: Three design sessions were conducted to inform the iteration of a prevention-focused, personalized mHealth, information-seeking app. The research team led sessions with 2 community member design teams. Design team participants described their goals, motives, and interests regarding prevention information using different approaches, such as collage and card sorting (design session 1), interaction with the app prototype (design session 2), and rating of cultural appropriateness strategies (design session 3).

Results: Each design team had 5 to 6 participants: 2 to 3 male participants and 3 female participants aged between 30 and 76 years. Design team participants shared their likes and dislikes about the sessions and the overall experience of the design sessions. Both African American and Hispanic teams reported positive participation experience. The primary reasons included the opportunity for their views to be heard, collectively working together in the design process, having their apprehension about mHealth reduced, and an opportunity to increase their knowledge of how they could manage their health through mHealth. The feedback from each session informed the following design sessions and a community-engaged process. In addition, the specific findings for each design session informed the design of the app for both communities.

Conclusions: This multi-method participatory research process revealed 4 key lessons learned and recommendations for future research in mHealth app design for African Americans and Hispanics. Lesson 1—community partnerships are key because they provide the *chain of trust* that helps African American and Hispanic participants feel comfortable participating in app research. Lesson 2—community-based participatory research principles continue to yield promising results to engage these populations in mHealth research. Lesson 3—interactive design sessions uncover participants' needs and development opportunities for mHealth tools. Lesson 4—multiple design sessions with different methods provide an in-depth understanding of participants' mHealth

XSL·FO

RenderX

preferences and needs. Future developers should consider these methods and lessons to ensure health apps in the marketplace contribute to eliminating health disparities and achieving health equity.

KEYWORDS

Introduction

Background

Nationally representative data from the Pew Research Center survey and the Health Information National Trends Survey show that Hispanic and African American adults are more likely than their White counterparts to use smartphones to access health information [1-3]. Hispanic and African American adults use their smartphones to access the internet to search for health information, communicate with health care providers, manage medicines, and use decision support tools, thereby reducing health gaps [1-3]. Despite these new digital options for prevention and management, Hispanic and African American populations are more vulnerable to some of the most common and preventable causes of illness and disease, such as diabetes and heart disease [4,5]. A well-designed, prevention-focused mobile health (mHealth) app that provides credible, culturally appropriate, and easy-to-understand information and recommendations could help better inform these groups, increase information-seeking, and aid decision-making to help prevent or delay chronic diseases. Previous research suggests that intended app users will be more likely to use and benefit from a prevention app if they are informed by their own lived experiences as well as being based on theoretical and methodological approaches that explain health behaviors [6,7]. Such apps can provide examples to mHealth app developers, public health practitioners, and researchers who want to ensure these products help eliminate health disparities, address communication inequities, and achieve health equity which is as follows [8]:

> *Is achieved when every person has the opportunity to 'attain his or her full health potential' and no one is 'disadvantaged from achieving this potential' because of social position or other socially determined circumstances.*

Limited health literacy may be a major barrier to mHealth app use, specifically among certain populations [9,10]. Healthy People 2030, the United States 10-year health objectives, define health literacy in both personal and organizational terms, and encompass not only comprehension of information but also seeking information and using it for decisions and actions [11]. According to the only nationally representative health literacy study of English-speaking adults in the United States, Hispanics and African Americans had lower average adequate health literacy than their White counterparts [11,12]. It is also important to consider how language is associated with health literacy in relation to the Hispanic population that prefers to speak Spanish or has limited English proficiency for health-related concerns. The National Assessment of Adult Literacy study also revealed that individuals who spoke only Spanish or a language other

than English had lower averages of health literacy skills than those whose primary language was English [12]. As the Hispanic population has grown substantially over the past decade, it is imperative that apps are available in Spanish to address their health information needs [13].

Participatory research methods are used in several disciplines, and in mHealth development, these methods have been shown to have positive effects when intended users are part of app development and testing processes [6,14]. This method involves the intended users in the research design and implementation of the intervention, product, or program [15]. Methods to engage intended users in app design and development, such as user-centered design, are known and published in the trade and academic literatures [16-18]. User-centered methods can inform all stages of app development and help refine and update apps as they mature. These methodologies provide insights into the motivations and challenges that users face and can help reveal any special circumstances or requirements that racial and ethnic groups, such as African American and Hispanic adults, may have with mHealth app use when seeking health information to manage their health.

Participatory approaches are imperative to inform and guide research with marginalized communities that have experienced unethical research practices and to engage these communities in research that is not part of their everyday lived experiences. Community-based participatory research (CBPR) has been effective in engaging racial and ethnic groups as well as marginalized communities in the development of mHealth tools. Its success in mHealth is rooted in researchers establishing trust with community partners while collaboratively working on the product development and evaluation processes of the intended product [6,14,19,20]. This is critical when engaging communities who are historically and understandably suspicious of or reluctant to engage in research [21-24].

Objectives

This methodological paper aims to describe a 1-year multi-method participatory research process that directly engaged English-speaking African American and bilingual or Spanish-speaking Hispanic adults in designing a prevention-focused, personalized mHealth, information-seeking smartphone app. The research reported in this paper was phase 1 of a 4-year process to iteratively refine the field test and revise an app. The study is being conducted by a multidisciplinary team of health literacy, communication, health services, public health, and computer science researchers. The process and app are grounded in CBPR, user-centered design principles, and health literacy techniques. This paper describes (1) the participatory approach, (2) the design session process, (3) participant-reported experiences of the design sessions, and (4)

recommendations and lessons learned for future research in mHealth app design for African Americans, Hispanics, and other racial and ethnic groups with a disproportionate burden of health disparities. This paper provides new information about how to combine participatory methods from different intellectual traditions to better discover African American and Hispanic adult health app needs through co-design.

Methods

App Design Session Theoretical Underpinnings and Methodological Approach

CBPR principles, user-centered design, and health literacy techniques were our core methods. Understanding the challenges in finding African American and Hispanic adults who might be interested in a research study on an mHealth app, the research team applied several CBPR principles. These principles included (1) building on strengths and resources within the community, (2) promoting a colearning and empowering process that attends to social inequalities, and (3) disseminating findings and knowledge to the user [25]. The research team collaborated with 2 key community partners, a National Institutes of Health–designated Research Center of Excellence in Race, Ethnicity, and Health Disparities Research, and a Hispanic-serving community-based organization. These partners are part of a *chain of trust* in their local Hispanic and African American communities because of the engagement infrastructures they have built over time. By affiliating with these partners, we became part of the chain. Collaboration with the community partners included routine meetings to seek input in the recruitment and retention strategies, content of marketing materials, and input for the development of the design sessions to ensure an equitable participatory research learning process for the participants. In addition, the participant design sessions supported a *co-learning and empowering process that attended to social inequalities* by establishing a bidirectional communicative process between the participants and research team to learn how the communities' needs can be better served. The research team practiced *disseminating findings and knowledge to the user* by continuously maintaining transparency in the research process and results throughout the study. To put the principles into practice for the app, the team applied user-centered and health literacy techniques.

Not only is a lack of community engagement a weakness in mHealth development but also many tools have been developed without a strong theoretical foundation. Our overall study combined the consumer information process (CIP) model and adult learning theory to inform information-seeking motivation and action [26,27]. The CIP and adult learning theories consider what motivates and engages adults in information-seeking and action. Adult learning theory posits that adults will invest in learning when they perceive a strong *need to know*. The theory's key ideas about the importance of adults' previous experience and their developmental stage were applied to inform app design and content. The CIP model's core concept is *individual processing capacity*, and the model's core assumption is that individuals are limited and intentional or goal-directed when they seek information and how much information they process [26]. These tenets align with health literacy research and its basic insight that people prefer relevant, easy-to-understand information that does not overwhelm them.

The CIP and adult learning theories and health literacy insights informed the choice of user-centered methods and the design session activities so that the team could develop an understanding of the motivations and interests of the participants in relation to health information–seeking through a prevention-focused mHealth app.

Background Work on an App Prototype Before the Design Sessions

Before receiving funding for the 4-year study and implementing the design sessions, the research team conducted a 6-month prefunding phase (year 0) that involved vetting the app concept and developing and programming an early-stage prototype. The core app content is from healthfinder.gov, an award-winning, federal consumer health information website available in English and Spanish that was designed based on health literacy principles [28]. The research team presented this to a community advisory group, the Maryland Community Research Advisory Board (MD-CRAB), to validate the app concept, receive general feedback on the prototype, and determine what amendments should be made before applying for research funding to support the phases shown in Figure 1. The MD-CRAB is a research advisory group established by the Maryland Center for Health Equity (M-CHE) at the University of Maryland School of Public Health. The MD-CRAB includes 22 members, most from the local African American and Hispanic communities, and they provide community insight and guidance to strengthen research and ensure the results of research benefit vulnerable populations, especially in African American and Hispanic populations.

Figure 1. Display of the 4-year smartphone health app research study phases.
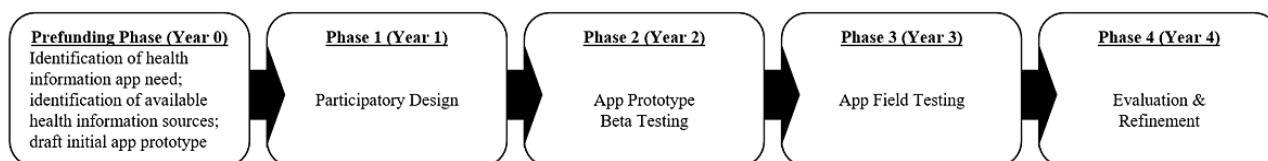
The African American and Hispanic MD-CRAB members reported that they believed people such as themselves might be willing to use an app that provided (1) information specific to their medical concern, (2) an array of health information, (3)

bilingual options, and (4) culturally tailored health information. We used this feedback to iterate the prototype and plan the research funding proposal. Once the study was funded in August

2018, we planned and began implementing phase 1 design sessions.

Design Session Participant Recruitment

Between September 2018 and December 2018, the research team worked with 2 community partner organizations in Prince George's and Montgomery Counties, Maryland, United States, to recruit English-speaking African American and bilingual or Spanish-speaking Hispanic adults. The organizations advised us on culturally tailored recruitment fliers. M-CHE led the African American recruitment, and the M-CHE Director asked the local barbers and stylists in the research community network in Prince George's County, Maryland, United States, to share the fliers. Requests to share the fliers with church members were made to local African American churches. An African American research team member presented the project at local church meetings to recruit participants. The research team also tried recruiting at local, free, well-attended health services events by many African American and Hispanic community members. Community health workers at Community Health and Empowerment through Education and Research (CHEER), a nonprofit service provider for Hispanic residents in the Takoma Park, Maryland neighborhood, distributed fliers and asked their clients to volunteer for the study. Interested African American and Hispanic community members called the project telephone number and left a message for a bilingual research assistant who called back and screened callers in English or Spanish to see if they met the study criteria. The goal was to recruit 1 group of African American adults and 1 group of bilingual or Spanish-speaking Hispanic adults who would commit to attending multiple design sessions over several months.

We screened participants based on race and ethnicity, age, sex, education, smartphone access, information-seeking habits, and willingness to participate in multiple sessions. Participants had to self-identify as African American or Hispanic or Latino and aged ≥18 years. We aimed for an equal distribution of self-identified male and female participants. As healthfinder.gov, the app's core content, is designed to be useful for adults with low health literacy, we were interested in working with community members with low health literacy and low health information–seeking habits. We screened participants to identify those who had no more than 2 years of community or technical college, had a smartphone or were willing to learn to use one, and answered questions that indicated they were a low to medium health information seeker and had low health literacy based on questions from the Health Information National Trends Survey [29]. Those who met the recruitment criteria were scheduled for the initial African American or Hispanic design sessions.

Design Session Planning

The research team wrote the moderators' guides in English that the bilingual research assistants translated into Spanish. The guides provided informed consent, followed by open-ended questions and tasks specific to the design sessions. The purpose of the sessions was to confirm the basic structure of the prototype app and solicit intended user ideas and feedback for

a revised app with additional content, features, and functions. We asked the same questions of the African American adults (in English) and Hispanic adults (in Spanish). A sample moderator's guide in English is available for review in the Multimedia Appendix 1. We scheduled the African American Study Director to lead the African American sessions and either a CHEER community health worker or a Hispanic bilingual research assistant to lead the Hispanic sessions in Spanish. We audio recorded each session and scheduled at least one bilingual or English-speaking research assistant to take notes during each session, depending on the session (English or Spanish). The design sessions were held on Saturdays at a local church for African American participants and evenings at the CHEER offices, a Hispanic-serving community-based organization, to make them convenient for participants based on their availability and preferences. Building trust and comfort between ourselves and participants was why we chose to hold the 3 sessions with the same 2 groups of participants. Participants were offered a gift card for US $25 for each session attended and an additional remuneration of US $25 if they completed all 3 sessions.

Ethics Approval

Each design session received approval from the University of Maryland Institutional Review Board (approval numbers 1292902-1, 1388156-1, and 1430335-1), and all project team members completed the Collaborative Institutional Training Initiative training for human subjects research before engaging with participants.

Design Session Procedures

User-center design methods engage the intended users of a product or service at different stages of design and testing processes to ensure maximum usability [30,31]. Our community partner organizations told us that although many community members have smartphones, they might not know much about how to download and use apps. In addition, we did not know how comfortable participants would be sharing information about their health experiences, beliefs, and technology use. As we were also applying CBPR and health literacy principles in how we engaged with participants, we chose methods to allow participants to draw on what was familiar to them and become comfortable talking about health topics and smartphone apps. We used hands-on activities and plain language in our study materials and design discussions to help participants fully understand and share the design process. To minimize any literacy challenges, we read aloud written materials, such as informed consent materials, written directions, or information on screens. During the 3 design sessions, we wanted participants to think broadly about their health and describe their goals, motives, and interests regarding prevention information using different approaches: (1) collage and card sorting (design session 1), (2) interaction with the app prototype (design session 2), and (3) rating of cultural appropriateness strategies (design session 3; Figure 2). In addition, participants could share their likes and dislikes about the sessions and the overall experience of the design sessions. We followed the same process as our separate African American and Hispanic groups for each design session.

XSL·FO

RenderX

Figure 2. Design session activities.
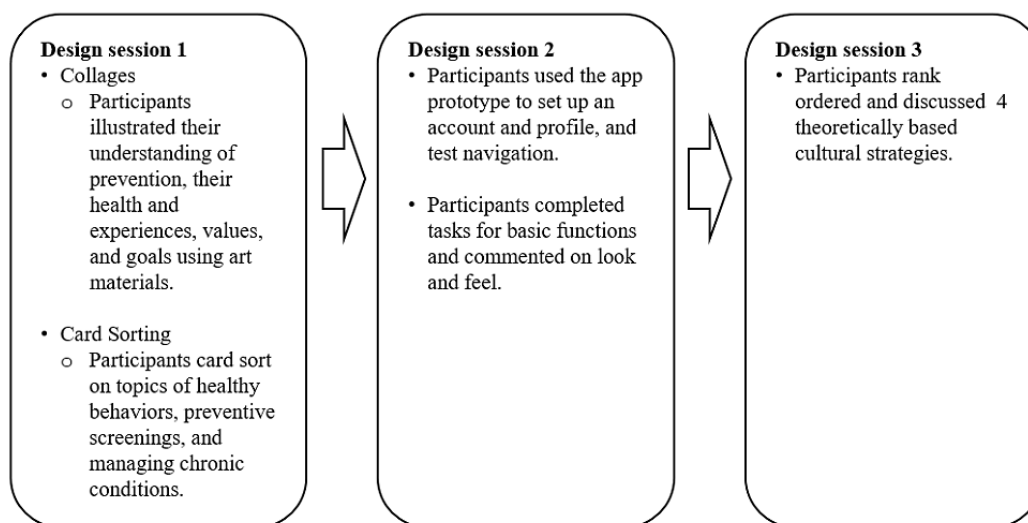
Design Session 1 Process

The goal of design session 1 was to learn participants' meanings and interpretations of health, prevention, and management, and the collage and card sorting activities involved participants creatively expressing their ideas, thoughts, interests, and desires about health generally and their health specifically. Collages are a *make tool* for people to create artifacts and express unspoken feelings and emotional states [32]. Participants were provided with tools such as poster board, colored pencils, markers, and printed images of people and activities that allowed them to create collages to explain what prevention meant to them, describe their present health, express their health values, and what they desired for their health in the future. Each participant had 30 minutes to create their own collage, and then the group spent 30 minutes on a *walking tour* so that each participant could tell the story of their collage to the group. The 1-hour card sorting activity allowed participants to have a tangible task to discuss prevention topics, which can be abstract for many people, and took advantage of in-depth group discussion to learn about participants' thoughts about the cards presented to them [33]. The card sorting activity involved each participant individually receiving a card stack on a group of health topics. The topic groups were healthy behaviors such as physical activity and healthy eating; preventive screenings and vaccinations, such as blood pressure checks or annual flu shots; and managing chronic conditions, such as chronic back pain or diabetes. We asked participants if they had questions about the terms on the cards and to talk aloud for 10 to 15 minutes about what the topics meant to them.

At the end of the discussion, we asked them to choose 2 to 3 cards that were most relevant to them. We repeated this process for round 2 with preventive screenings and vaccinations. For round 3, we asked them to look at 5 chronic disease cards and tell us the first thing they thought of. At the end of the session, we collected the collages and stacks of the sorted cards for analysis. The research assistants summarized the recordings and their session notes and organized responses by discussion

questions in spreadsheets. The research team members who attended the session, which included the principal investigator (PI) and the study director, debriefed, examined the collages for similarities and differences, and reviewed participants' comments by question. The PI and study director discussed the collages and participants' comments and recommended app changes to the full research team, who collectively decided what to revise before design session 2.

Design Session 2 Process

The goal of design session 2 was to learn how participants navigated the app prototype, and the session involved participants interacting with the revised prototype app and providing input on the interface, functions, and overall look and feel. A key step in user-centered design is to provide prototypes to a small number of intended users to try real-world tasks and scenarios [30]. Our participant groups included people with vision or literacy challenges; so, at the session sites we set up full-size computer monitors and connected a smartphone with the app to display the app screens for easier readability. The participant, research team facilitator, and research assistant sat around the monitor, and then the participant held the smartphone to complete the tasks while being observed. We read aloud the directions and information. We video recorded each session, and the research assistant took notes. Participants had 4 tasks to complete: (1) review the introductory content of the app that describes the app purpose, (2) perform a nonpersonalized health information search in the app, (3) perform a basic personalized health information search in the app entering data about their age, sex, and pregnancy status, and (4) create a personalized profile with demographic and health information data. We asked them to talk aloud while they did the tasks and explain what they were doing and thinking. The research assistants summarized the recordings and notes of participants and questions in spreadsheets. The PI and the study director, who were present during the testing, reviewed the responses and recommended app changes to the full research team for discussion and implementation decisions.

Design Session 3 Process

Knowing and including culturally appropriate references is a key tenant of effective health communication practice, which can also be applied to digital health tools [34,35]. The goal of design session 3 was to learn design team participants' cultural preferences and needs in relation to health, and we explored culturally appropriate strategies reported in the literature that are necessary for relevant and useful health information and messages [34,36]. On the basis of comments and suggestions from design sessions 1 and 2 as well as the literature on culturally tailored strategies, the research team developed a packet of activities that asked participants to evaluate the different strategies: (1) cultural appropriateness appearance, (2) evidence of health issues specific to African Americans and Hispanics, (3) specific language and linguistic content, and (4) sociocultural values and characteristics. For the appearance evaluation task, for example, the packet included possible images for the app, and for evidence, examples of health statistics relevant to either African American or Hispanic adults. The research team discussed each strategy with participants. This session also involved participants viewing and ranking their interest in web content and tools, such as a health services locator, nutrition planner, medication information source, and an air quality tracker, that the research team was considering for the final app. These activities provided further understanding of what the participants would like to integrate in the app to increase and motivate engagement.

The research assistants summarized the audio recordings and their written notes on the spreadsheets. The PI and the study director summarized the discussions for the full research team, and the team reviewed participants' app content addition preferences. The research team convened the design groups in a separate final community member report-back meeting. The research team shared the revised app with the design session participants so they could see how their input manifested itself in the app. This was a *report-back* and a user validation session, not data collection, to close the loop with our community co-designers. The authors took notes but did not audio record the final session.

Process for How Design Session Data Informed the App

Each session was audio or video recorded, and sometimes both depending on the type of session, and transcribed and the

Spanish-speaking sessions were translated into English by bilingual research assistants on the team. Data collected from each session were entered into spreadsheets and categorized based on the design session discussion guide questions. The input from each design session informed the app development and was organized according to the (1) changes to the existing prototype to make it more attractive, easier to use, and relevant; (2) missing features that participants felt would motivate their app use and continued engagement; (3) suggestions on how to present and explain the importance of preventing illness and disease; and (4) suggestions on how to explain the personalization features, such as personal and family health history information and goal-setting. The analysis of the data from each design session is not included in this paper, as the intent of this methodological paper is to propose and discuss the multi-method co-design process we used. This paper provides the feedback from the design session team participants' on their experiences participating in the overall design process. Participant feedback was provided informally during each design session.

Results

Design Session Recruitment and Attendance

The M-CHE recruitment efforts at one local African American church resulted in 6 English-speaking African American adults, and the CHEER outreach produced 5 bilingual or Spanish-speaking Hispanic adults participating in the sessions (Table 1). The racial, ethnic, and sex distributions were 3 African American women, 3 African American men, 3 Hispanic women, and 2 Hispanic men. Participants' were aged between 30 and 76 years. Three design sessions and a report-back in English for African American participants and 3 design sessions and a report-back in Spanish for Hispanic participants occurred between December 2018 and May 2019. The groups were conducted separately to allow each group to communicate in their preferred language. Each design session lasted 2 to 3 hours. Although we held the sessions at the local church or CHEER office, participant attendance varied by session, sometimes resulting in 5 to 6 participants attending.

Table 1. Design session participant demographics (n=11).

Demographics	African American participants (n=6)	Hispanic participants (n=5)
Age (years), range	30-76	35-76
Sex, n (%)		
Male	3 (50)	2 (40)
Female	3 (50)	3 (60)
Total number of participants (design session 1), n (%)	6 (100)	5 (100)
Total number of participants (design session 2), n (%)	5 (84)	5 (100)
Total number of participants (design session 3), n (%)	5 (84)	5 (100)

Design Session Participant Experiences

Overview

Both African American and Hispanic design team members reported a positive participation experience. The primary reasons for their experiences included (1) the opportunity for their views to be heard, (2) collectively working together in the design process, (3) having their apprehension about mHealth reduced, and (4) an opportunity to increase their knowledge of how they could manage their health through mHealth. This feedback informed us about how we approached each design session and how we engaged the participants. In addition, the specific findings for each design session (not reported in this methodology paper) informed the design of the app for both communities.

Reason 1: Opportunity for Their Views to Be Heard

Both African American and Hispanic participants said that having their *perspectives* heard was important. Hispanic participants reported the following:

> I like that you take our views and perspectives into account as elders and facilitate access to cell phone research.

> I appreciate very much that you are taking all of us into account.

Reason 2: Collectively Working Together in the Design Process

African American participants reported the following: "I agree with everyone but I think the gathering, us networking, being able to share our experiences, being able to talk about ourselves openly and not being judgmental about what somebody else is going through."

This quote also reflected another reason for participants' positive experiences, as it reflects their appreciation of working together in the design process. Participants felt valued and respected, resulting in comfortably sharing their honest opinions with one another and the research team about what they thought of the app. Another participant reported the following: "In regards to the app I think that it is actually a good tool that we are part of and that you guys are trying to put out there for people to have more knowledge and have more control of their health."

In addition, an African American participant inquired how the research team established a relationship with the community to conduct this research, and we explained how the community partnerships connected us to the participants. The participant responded by saying the following:

> I think that this is a good thing because we really don't have enough of that [research] and with so much going on...we need to do research...

Reason 3: Having Their Apprehension About mHealth Reduced

The third reason for the participants' positive experience with using the app was the reduced apprehension about mHealth. An African American participant reported the following:

> You guys explaining everything to me. My vision is bad and I suffer sometimes when I'm looking at things. But you guys made it easy for me by assisting me. I've kind of understood and I'm not lost. At first I was concerned because I thought, 'Oh gosh, I'm going to sit here and be lost. I'm not going to be able to see this...etc.' But you guys made sure that I'm comfortable, I'm not intimidated...

This particular participant reported apprehension while using mHealth technologies but wanted to be involved in the study to learn how to use the technology and overcome their fear. Their comment also provided insight into the app function that needed to be adjusted based on physiological abilities.

Reason 4: An Opportunity to Increase Their Knowledge of How They Could Manage Their Health Through mHealth

Finally, the other positive experience participants reported was an opportunity to increase their knowledge of how they could manage their health through mHealth. A Hispanic participant stated the following:

> I like what they (the research team) are doing (the session) at the University of Maryland for the Hispanic community. There are many times that they (Hispanic community) don't find or don't know that a stomach ache or something simple can become something chronic. I like it [the app] a lot.

African American participants reported the following:

> That's what I got out of it. The app enables you to take care of yourself.

Discussion

Principal Findings

Overview

Our study design aimed to create a positive app development experience for a small group of English-speaking African American and Spanish-speaking Hispanic adults and generate rich data to inform an app development project. The mHealth literature reports the importance of community participation and user-centered design, and participants' responses to the research methods reported in this paper confirm these strategies. However, this study went further than others in applying health literacy principles at every step of the mHealth design process—from recruitment of participants to selection of app content, feature design, and the report-back process. Our study provides 4 key lessons and recommendations in mHealth Design for racial and ethnic groups with health disparities [37,38].

Lesson 1: Community Partnerships Are Key in Engaging Racial and Ethnic Groups With a Disproportionate Burden of Health Disparities in mHealth App Development Research

The research team in collaboration with the community partner organizations was able to successfully recruit and retain the necessary number of participants over a 6-month design session series. We tapped into the *chain of trust* that our community

partners had with the intended app users. The research team remained connected with the community partner organizations to discuss strategies to engage Hispanics and African Americans in the research process and app design. For example, when the research team and community partners were initially designing recruitment materials (ie, fliers and recruitment screening tools), both community partners stated that one standard flier for both groups would not be effective. On the basis of their experience with the local community members, the African American flier required a message stating how their community would benefit from the technology if they participated. The community partner stated that African Americans are constantly bombarded with negative statistics and information about their community, and they want to hear information framed as gains rather than losses (eg, what we will gain if we participate). The Hispanic community partner advised us to frame their recruitment message as both the gains and losses of healthier lifestyle adoption (eg, our community can gain new information and will lose out if we do not participate). Similar input informed design session moderator guides and session materials. Discussions with community partners also revealed similarities in how African American and Hispanic participants would prefer to engage in design sessions as it relates to the type of materials used, topics discussed, and technology preferences. The constant feedback and support from the community partners helped create an engaging design session environment that produced the positive and transparent participant experiences reported. We were able to apply cultural tailoring not only to the app but also in the development of the design sessions. Routine interaction and feedback with the community partners assisted us in navigating and avoiding potential pitfalls that might have occurred had we not established the collaboration.

Lesson 2: Application of CBPR Principles Continues to Yield Promising Results When Engaging Racial and Ethnic Groups With a Disproportionate Burden of Health Disparities in mHealth Research

Owing to the equitable and colearning environment that CBPR brings to the research environment, people are more likely to see value in participating in the research process because they see how their input is valued while also gaining additional information and resources. Several participants stated they appreciated that they had the opportunity to share their input about a product intended for them. A large body of literature demonstrates the benefits of applying some or all of the CBPR principles when engaging community members in research [39-41]. CBPR is also being applied specifically in mHealth app development and encouraged by other mHealth and public health researchers and developers [6,14,42,43].

Lesson 3: Interactive Design Sessions Allow Individuals to Uncover Their Needs and Opportunities for the mHealth Tool Being Developed

Several of the design session participants reported that participating in the sessions increased their awareness of health issues and the need for an mHealth information-seeking app. When presented with the prototype app for review, several reported not knowing that this information was or could be available to them and also stated how helpful it would be for themselves and others they know. Adult learning theory posits that adults will invest in learning when they perceive a strong *need to know*, and an mHealth information-seeking app can be the prompt. The team's design approach provides a tangible method that allows participants to uncover their needs and opportunities for the mHealth tool. It allows participants to manipulate something and discover something new in the process.

Lesson 4: Sustained Design Sessions Using Multiple Approaches Can Provide an In-depth Understanding in mHealth Preferences and Needs in Appearance, Function, and Content

Participants had the opportunity to provide their views about health and its relationship to information-seeking through various methods. These types of approaches aid developers and help them avoid unnecessary and potentially biased errors in the development process [6]. The use of design sessions that built on each other (eg, collaging, card sorting, user–app interaction, and cultural appropriateness strategies) allowed the research team and participants to learn about each other and develop a deeper understanding of what motivates and maintains engagement in mHealth apps among health disparity populations.

Strengths and Limitations

Two of our study's strengths focus on health literacy issues for smartphone app design and language access for people who prefer health information in languages other than English. Excluding higher education and information-seeking participants extended recruitment because many people who expressed interest did not meet these 2 criteria. However, we felt it was important to work with participants with limited health literacy and design an app for their needs that would likely work well for others with higher health literacy [28]. Working with English- and Spanish-speaking groups in parallel allowed us to see cultural and language similarities and differences in how groups might vary in app use, although we did not set the goal of conducting an explicit comparative study.

The extended engagement over 4 sessions with the same group of African American and Hispanic adults is also a strength of our study. We were able to use multiple methods to collect participant insights on different app design issues, and our participants informed our decisions at different development stages and experienced app development along with the research team. Furthermore, our app developers attended many design sessions, and they could visualize and hear participants in their own words, which kept them grounded in our participants' lived experiences.

Having participants participate in real-time app development also had its limitations. They could not experience the app in its fully functional form as we were learning and prototyping based on their input. The research team continued to iterate the app for several months after the design sessions concluded. One key feature of the app, a personalization algorithm, was added during the final development phase. We described the algorithm and personalization of the participants, but they could not try it in real time.

Although our design groups had a typical number of participants (5 to 6) for these types of user-centered design studies, a limitation is the small number of participants, as they cannot represent the needs, experiences, and perspectives of all African American and Hispanic smartphone users. However, the multiple design sessions with the same participants performing concrete tasks allowed them to provide detailed information not available through other methods. As other researchers note, user-centered methods have high information yields and rich data [37].

Conclusions

Our study contributes to the small and growing literature on the involvement of marginalized groups in mHealth design and evaluation. Although health literacy and language access issues are cited as barriers to some groups' engagement with mHealth, our project shows how careful attention to these issues can be incorporated into standard user-centered methods and CBPR principles. Participants' comments about how much they appreciated the sessions and the chance to engage with mHealth tools on their own terms validates our approach.

To reduce cultural bias errors, sustain engagement, and provide culturally relevant mHealth information apps to African American and Hispanic adults who continue to be affected disproportionately by preventable chronic conditions, developers should consider a multifaceted participatory research process that includes user-centered design and health literacy approaches. Because of social factors that increase their vulnerability to chronic diseases, it is imperative that these communities have access to digital tools that truly address their health information needs. Although health information–seeking apps are not a common or popular type of app, they have significant potential to serve as a tool that can support individuals' health care management. The comments from our participants indicate that these communities would greatly benefit from an mHealth information-seeking tool that is personalized to their health information needs. Future developers should consider more integrative app development approaches to ensure that the apps on the market contribute to eliminating health disparities and achieving health equity. It is the responsibility of those who are developing mHealth apps to ensure that studies are inclusive of the communities they intend to serve.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sample moderator guide in English for a design session (session 1 of 3 sessions).
[DOCX File , 20 KB - formative_v6i3e26764_app1.docx]

References

1. Bhuyan SS, Lu N, Chandak A, Kim H, Wyant D, Bhatt J, et al. Use of mobile health applications for health-seeking behavior among US adults. J Med Syst 2016 Jun;40(6):153. [doi: 10.1007/s10916-016-0492-7] [Medline: 27147516]
2. Smartphones help Blacks, Hispanics close digital gap with Whites. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2017/08/31/smartphones-help-blacks-hispanics-bridge-some-but-not-all-digital-gaps-with-whites/ [accessed 2021-12-24]
3. Demographics of mobile device ownership and adoption in the United States. Qure4U. URL: https://www.qure4u.com/demographics-of-mobile-device-ownership-and-adoption-in-the-united-states/ [accessed 2021-12-24]
4. Hispanics' health in the United States. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/media/releases/2015/p0505-hispanic-health.html [accessed 2021-12-24]
5. African American health. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/vitalsigns/aahealth/index.html [accessed 2021-12-24]
6. Brewer LC, Fortuna KL, Jones C, Walker R, Hayes SN, Patten CA, et al. Back to the future: achieving health equity through health informatics and digital health. JMIR Mhealth Uhealth 2020 Jan 14;8(1):e14512. [doi: 10.2196/14512] [Medline: 31934874]
7. Vollmer Dahlke D, Fair K, Hong YA, Beaudoin CE, Pulczinski J, Ory MG. Apps seeking theories: results of a study on the use of health behavior change theories in cancer survivorship mobile apps. JMIR Mhealth Uhealth 2015 Mar 27;3(1):e31 [FREE Full text] [doi: 10.2196/mhealth.3861] [Medline: 25830810]

8. Health equity. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/chronicdisease/healthequity/index.htm [accessed 2021-12-24]

9. Kreps GL. The relevance of health literacy to mHealth. Stud Health Technol Inform 2017;240:347-355. [Medline: 28972527]

10. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]

11. Health literacy in healthy people. health.gov. URL: https://health.gov/our-work/national-health-initiatives/healthy-people/healthy-people-2030/health-literacy-healthy-people-2030 [accessed 2021-12-24]

12. The health literacy of America's adults: results from the 2003 national assessment of adult literacy. National Center for Education Statistics. 2006. URL: https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006483 [accessed 2021-12-24]

13. U.S. Hispanic population surpassed 60 million in 2019, but growth has slowed. Pew Research Center. 2020. URL: https://www.pewresearch.org/fact-tank/2020/07/07/u-s-hispanic-population-surpassed-60-million-in-2019-but-growth-has-slowed/ [accessed 2021-12-24]

14. Brewer LC, Hayes SN, Caron AR, Derby DA, Breutzman NS, Wicks A, et al. Promoting cardiovascular health and wellness among African-Americans: community participatory approach to design an innovative mobile-health intervention. PLoS One 2019;14(8):e0218724 [FREE Full text] [doi: 10.1371/journal.pone.0218724] [Medline: 31430294]

15. Cornwall A, Jewkes R. What is participatory research? Soc Sci Med 1995 Dec;41(12):1667-1676. [doi: 10.1016/0277-9536(95)00127-s]

16. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]

17. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]

18. Norman DA. The Design of Everyday Things. Cambridge, MA: MIT Press; 1998.

19. de Peralta M, Smithwick J, Torres M. Perceptions and determinants of partnership trust in the context of community-based participatory research. J Health Disparities Res Pract 2020;13(1):67-95 [FREE Full text]

20. Jagosh J, Bush PL, Salsberg J, Macaulay AC, Greenhalgh T, Wong G, et al. A realist evaluation of community-based participatory research: partnership synergy, trust building and related ripple effects. BMC Public Health 2015 Jul 30;15(1):725 [FREE Full text] [doi: 10.1186/s12889-015-1949-1] [Medline: 26223523]

21. Corbie-Smith G, Thomas SB, St George DM. Distrust, race, and research. Arch Intern Med 2002 Nov 25;162(21):2458-2463. [doi: 10.1001/archinte.162.21.2458] [Medline: 12437405]

22. George S, Duran N, Norris K. A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. Am J Public Health 2014 Feb;104(2):e16-e31. [doi: 10.2105/AJPH.2013.301706] [Medline: 24328648]

23. Simmonds G. African American participation in public health research. ABNF J 2008;19(2):69-72. [Medline: 18494406]

24. Scharff DP, Mathews KJ, Jackson P, Hoffsuemmer J, Martin E, Edwards D. More than Tuskegee: understanding mistrust about research participation. J Health Care Poor Underserved 2010 Aug;21(3):879-897 [FREE Full text] [doi: 10.1353/hpu.0.0323] [Medline: 20693733]

25. Israel BA, Schulz AJ, Parker EA, Becker AB, Community-Campus Partnerships for Health. Community-based participatory research: policy recommendations for promoting a partnership approach in health research. Educ Health (Abingdon) 2001;14(2):182-197. [doi: 10.1080/13576280110051055] [Medline: 14742017]

26. Ross I, Bettman JR. An information processing theory of consumer choice. J Mark 1979;43(3):124. [doi: 10.2307/1250155]

27. Knowles MS, Holton III EF, Swanson RA. The Adult Learner: The Definitive Classic in Adult Education and Human Resource Development. London: Routledge; 2011.

28. Health literacy online: a guide for simplifying the user experience. Office of Disease Prevention and Health Promotion. URL: https://health.gov/healthliteracyonline/ [accessed 2021-12-24]

29. What is HINTS? Health Information National Trends Survey. URL: https://hints.cancer.gov/ [accessed 2021-12-24]

30. Designing the User Interface, new 6th Edition. HCIL Archive. URL: http://www.cs.umd.edu/hcil/DTUI6/ [accessed 2021-12-24]

31. User experience basics. usability.gov. URL: https://www.usability.gov/what-and-why/user-experience.html [accessed 2021-12-24]

32. Sanders EBN. From user-centered to participatory design approaches. In: Design and the Social Sciences. Boca Raton, Florida: CRC Press; 2002.

33. Conrad LY, Tucker VM. Making it tangible: hybrid card sorting within qualitative interviews. J Documen 2018 Nov;75(4):397-416. [doi: 10.1108/jd-06-2018-0091]

34. Kreuter MW, Lukwago SN, Bucholtz DC, Clark EM, Sanders-Thompson V. Achieving cultural appropriateness in health promotion programs: targeted and tailored approaches. Health Educ Behav 2003 Apr;30(2):133-146. [doi: 10.1177/1090198102251021] [Medline: 12693519]

35. Tan NQ, Cho H. Cultural appropriateness in health communication: a review and a revised framework. J Health Commun 2019;24(5):492-502 [FREE Full text] [doi: 10.1080/10810730.2019.1620382] [Medline: 31132946]

36. Sanders Thompson VL, Cavazos-Rehg PA, Jupka K, Caito N, Gratzke J, Tate KY, et al. Evidential preferences: cultural appropriateness strategies in health communications. Health Educ Res 2008 Jun;23(3):549-559. [doi: 10.1093/her/cym029] [Medline: 17631608]

37. Derksen ME, Jaspers MW, van Strijp S, Fransen MP. Mobile health for smoking cessation among disadvantaged young women during and after pregnancy: user-centered design and usability study. JMIR Form Res 2021 Aug 04;5(8):e24112 [FREE Full text] [doi: 10.2196/24112] [Medline: 34346895]

38. Stowell E, Lyson M, Saksono H, Wurth RC, Jimison H, Pavel M, et al. Designing and evaluating mHealth interventions for vulnerable populations: a systematic review. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; Apr 21-26, 2018; Montreal QC Canada. [doi: 10.1145/3173574.3173589]

39. Wallerstein NB, Duran B. Using community-based participatory research to address health disparities. Health Promot Pract 2006 Jul;7(3):312-323. [doi: 10.1177/1524839906289376] [Medline: 16760238]

40. Wallerstein N, Duran B. Community-based participatory research contributions to intervention research: the intersection of science and practice to improve health equity. Am J Public Health 2010 Apr;100(S1):S40-S46. [doi: 10.2105/ajph.2009.184036]

41. Horowitz CR, Robinson M, Seifer S. Community-based participatory research from the margin to the mainstream. Circulation 2009 May 19;119(19):2633-2642. [doi: 10.1161/circulationaha.107.729863]

42. Smith SA, Whitehead MS, Sheats J, Mastromonico J, Yoo W, Coughlin SS. A community-engaged approach to developing a mobile cancer prevention app: the mCPA study protocol. JMIR Res Protoc 2016 Mar 02;5(1):e34 [FREE Full text] [doi: 10.2196/resprot.5290] [Medline: 26935995]

43. Unertl KM, Schaefbauer C, Campbell T, Senteio C, Siek KA, Bakken S, et al. Integrating community-based participatory research and informatics approaches to improve the engagement and health of underserved populations. J Am Med Inform Assoc 2016 Jan;23(1):60-73 [FREE Full text] [doi: 10.1093/jamia/ocv094] [Medline: 26228766]

Abbreviations

CBPR: Community-Based Participatory Research
CHEER: Community Health and Empowerment through Education and Research
CIP: consumer information process
M-CHE: Maryland Center for Health Equity
MD-CRAB: Maryland Community Research Advisory Board
PI: principal investigator

<u>Original Paper</u>

The Effect of the Imacoco Care Psychoeducation Website on Improving Psychological Distress Among Workers During the COVID-19 Pandemic: Randomized Controlled Trial

Kotaro Imamura[1], MPH, PhD; Natsu Sasaki[1], MD; Yuki Sekiya[1], PhD; Kazuhiro Watanabe[2], PhD; Asuka Sakuraya[3], MPH, PhD; Yutaka Matsuyama[4], PhD; Daisuke Nishi[1,5], MD; Norito Kawakami[1], MD

[1]Department of Mental Health, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[2]Department of Public Health, Kitasato University School of Medicine, Sagamihara, Japan

[3]Division of Public Health, Department of Hygiene and Public Health, School of Medicine, Tokyo Women's Medical University, Tokyo, Japan

[4]Department of Biostatistics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[5]Department of Public Mental Health Research, National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo, Japan

Corresponding Author:
Norito Kawakami, MD
Department of Mental Health
Graduate School of Medicine
The University of Tokyo
7-3-1, Hongo, Bunkyo-ku
Tokyo, 113-0033
Japan
Phone: 81 3 5841 3522
Fax: 81 3 5841 3392
Email: nkawakami@m.u-tokyo.ac.jp

Abstract

Background: The prolonged COVID-19 pandemic has affected mental health among workers. Psychoeducational intervention via a website could be effective for primary prevention of mental illness among workers in the current COVID-19 pandemic.

Objective: The aim of this randomized controlled trial is to examine the effect of a newly developed online psychoeducational website named Imacoco Care on reducing psychological distress and fear about COVID-19 infection among workers.

Methods: Participants in the study were recruited from registered members of a web survey company in Japan. Participants who fulfilled the eligibility criteria were randomly allocated to the intervention or control group. Participants in the intervention group were invited to access the Imacoco Care program within a month after the baseline survey. The Kessler Psychological Distress Scale (K6) and the Fear of COVID-19 Scale (FCV-19S) scores were obtained at baseline and at 1- and 3-month follow-ups.

Results: A total of 1200 workers were randomly allocated to the intervention and control groups (n=600 [50%] per group). The Imacoco Care intervention group showed a significant favorable effect on K6 scores (P=.03) with a small effect size (ES; Cohen d=–0.14) and an adverse effect on FCV-19S scores (P=.01) with a small ES (Cohen d=0.16) at 3-month follow-up. In the per protocol analysis (including only participants who had read the Imacoco Care content at least 1 time), the Imacoco Care intervention group also showed a significant favorable effect on reducing K6 scores (P=.03), while an adverse effect on FCV-19S scores was not significant (P=.06) in the intervention group at 3-month follow-up.

Conclusions: A web-based psychoeducation approach may be effective for improving psychological distress among workers; however, it may be important not only to distribute information but also to encourage active engagement with the content of the program to prevent adverse effects of psychoeducational intervention.

Trial Registration: University Hospital Medical Information Network Clinical Trials Registry (UMIN-CTR) UMIN000042556; https://upload.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000048548

XSL•FO
RenderX

KEYWORDS

COVID-19; education; internet-based intervention; occupational groups; psychological distress; mental health; digital health; health intervention; psychoeducation

Introduction

Background

The prolonged COVID-19 pandemic has affected the mental health not only of essential health care workers [1-5] but also of other workers (ie, non-health-care) [6,7]. Previous studies have reported a high prevalence of depression (34.6%), anxiety (42.3%), and psychological distress (65.1%) among workers [6,7], as well as a decline in the health-related quality of life [8] during the COVID-19 pandemic. Providing effective primary prevention stress management interventions is essential for protecting and promoting the mental health of workers during the prolonged COVID-19 pandemic [9].

Psychoeducational intervention is 1 of the effective treatments for primary prevention of mental health problems in the working population. A recent systematic review aiming to identify and summarize interventions for dealing with mental health issues of health care workers during infectious disease outbreaks, including COVID-19, reported that emotional and psychological intervention, including psychoeducation, is an important approach [10], especially since the 2000s, online-delivered psychoeducational interventions have received much attention. A recent meta-analysis reported that online psychological and psychoeducational interventions, which are mainly based on the principle of cognitive behavioral therapy (CBT), have a significant small effect (pooled standardized mean difference [SMD] –0.26) on reducing depressive symptoms in nondepressed and varied populations, with a moderate quality of evidence [11]. One of the effective online formats for delivering psychoeducational information is a website. Some previous studies before the COVID-19 pandemic have reported that psychoeducation websites significantly improved depressive symptoms among community residents with elevated depressive symptoms [12] and workers who visited a mental health specialist in the past month at baseline [13]. Psychoeducational intervention via a website could also be effective for primary prevention among workers in the current COVID-19 pandemic.

Although online/website-based psychoeducational intervention may be promising for improving the mental health of workers in the COVID-19 pandemic, only a few randomized controlled trials (RCTs) have been conducted during the COVID-19 pandemic. Regarding studies targeting community residents, 2 RCTs applying CBT-based intervention among people with elevated mental symptoms showed favorable effects on mental health outcomes [14,15], while another RCT applying low-intensity online mindfulness-based intervention in the general population could not find a significant effect on any outcomes [16]. In the working population, only 1 study focusing on health care workers who had provided direct face-to-face health care to patients with a diagnosis of infection with COVID-19 was conducted during the COVID-19 pandemic [17]. This RCT reported that a self-guided CBT and mindfulness-based mobile health intervention (the PsyCovidApp) significantly improved overall mental symptoms (total score of depression, anxiety, and stress) among health care workers who used psychotropic medications (SMD –0.29) or received psychotherapy (SMD –0.25) at baseline, while there was no significant intervention effect among the whole sample [17]. Developing an effective online psychoeducational intervention and examining its effect on promoting mental health are needed for workers during the COVID-19 pandemic. Although demonstrating the efficacy of such a cost-effective psychoeducation website would be beneficial, there is no other RCT aiming to examine the effect of a psychoeducational website on improving mental health problems in the working population in the current COVID-19 pandemic. Therefore, a further RCT is needed to examine the effect of psychoeducational intervention in a sample of workers.

Objective

The aims of this study were to examine the effect of a newly developed self-guided, low-intensity, and easy-to-disseminate psychoeducational website named Imacoco Care on improving mental health (ie, reducing psychological distress and fear about COVID-19 infection) among workers, using an RCT design.

The hypotheses in this study were as follows: (1) Imacoco Care would reduce psychological distress as the primary outcome among workers, and (2) Imacoco Care would reduce fear about COVID-19 as the secondary outcome among workers.

Methods

Trial Design

This study was an RCT. The allocation ratio of the intervention group to the control group was 1:1. The study protocol was registered at the University Hospital Medical Information Network Clinical Trials Registry (UMIN000042556). This paper was reported according to the Consolidated Standards of Reporting Trials (CONSORT) guidelines [18,19].

Ethical Considerations

The Research Ethics Review Board of the Graduate School of Medicine and Faculty of Medicine, University of Tokyo, approved the study procedures (number 3083-6).

Participants

Participants in this study were recruited from registered members of a web survey company in Japan (Rakuten Insight, Inc [20]). Over 2 million members are registered in this company. Of those, 10,000 adult workers who were potentially eligible for the study were randomly selected. These candidates received a URL of a webpage that included detailed information about the study. The aims and procedures of the study were fully explained on the webpage. Participants were asked to click the Agree button to show their consent to participate in the study; then they proceeded to the baseline questionnaire page. Written consent was not required by the National Ethical Guidelines for Epidemiologic Research, Japan; the Research

Ethics Review Board of the Graduate School of Medicine and Faculty of Medicine, University of Tokyo, approved this procedure for obtaining participants' consent.

Eligibility Criteria of This Study

The inclusion criteria of this study were (1) adults (over 20 years old) and (2) full-time employees. The exclusion criteria were (1) having 15 sick leave days or more during the past 3 months, (2) having consultations with mental health professionals during the past month, and (3) having participated in this intervention program in the past.

Intervention Program: Imacoco Care

Imacoco Care is a newly developed self-guided psychoeducation website-based program for providing information about coping with stress or problems under the conditions of the COVID-19 pandemic [21]. Imacoco Care provides information for managing stress and maintaining mental health among people staying at home under the state of emergency. The content is mainly composed of text, illustrations, video, and audio narration. There are about 30 pages in total, with around 1000 Japanese characters (500 words in English) per page. All content is explained in plain language, easy to understand for most workers, and does not presume prior knowledge of stress management.

Imacoco Care provides information about evidence-based psychological interventions for psychological distress (see Table 1), and the website can be accessed anywhere the internet is available. All of the Imacoco Care content was developed by the authors. Recent studies focusing on mental health among workers in the COVID-19 pandemic have reported that psychological/physical interventions, such as mindfulness, CBT-based psychoeducation, and physical activity, could improve mental health problems during the COVID-19 pandemic [14,22,23]. Details of the content included on the Imacoco Care website are listed in Table 1.

Table 1. Overview of the content of the Imacoco Care website.

Modules	Content
Mindfulness	This module provides information about the benefits of mindfulness and how to engage in it with easy-to-understand audio and video.
Behavioral activation (BA)	This module introduces how to relate behaviors/actions to feelings, how to maintain mental energy, and how to perform activities that lead to good feelings.
Physical activity	This module provides knowledge and tips for leading an active life while people are asked to stay at home, being properly aware of the spread of the new coronavirus.
Sleep education	This module provides information about how to regulate the sleep rhythm without using medication.
Tips for working from home	This module provides tips to make telecommuting (working from home) a little more comfortable.
Coping with the stress about COVID-19	This module provides stress management tips that can be practiced in daily life.

Mindfulness

An operational definition of mindfulness is "the awareness that emerges through paying attention on purpose, in the present moment, and nonjudgmentally to the unfolding of experience moment by moment" [24]. A previous meta-analysis reported that mindfulness meditation programs yielded moderate evidence of improved anxiety (effect size [ES]=0.38 at 8 weeks and 0.22 at 3-6 months) and depression (ES=0.30 at 8 weeks and 0.23 at 3-6 months) [25]. Imacoco Care provides information about the benefits of mindfulness and how to engage in it in the following lessons, with easy-to-understand audio and video: (1) For When You Feel Uncomfortable—an Introduction to Mindfulness; (2) The Body Relates to Our Feelings—Using mindfulness to Pay Attention to the Sensations of the Body; (3) Mindfulness by Narrating Like a Live Broadcast; (4) Mindful Eating; and (5) Self-Compassion Exercises to Calm Painful Feelings [26].

Behavioral Activation

Behavioral activation (BA), one of the most readily applied techniques in CBT, is a process to increase pleasurable and rewarding activities using behavioral strategies, such as activity scheduling [27]. A previous meta-analysis showed that CBT including BA could significantly reduce the risk of the onset of major depression (incidence rate ratio=0.62) [28]. Imacoco Care introduces how to relate behaviors/actions to feelings, how to maintain mental energy, and which activities lead to good feelings in the following lessons: (1) Tips for Keeping Your Mind Energized During Difficult Daily Life, (2) The Relationship Between What You Do or Don't Do and Your Feelings, (3) Reflecting on Your Life Patterns, (4) Think of a "Safe Activity" That Doesn't Fit Into the "Three Cs" (Crowded Places, Close Contact, Confined and Enclosed Spaces), (5) Let's Make Your Action Plan!, and (6) Reflections: Tips for Acting to Keep Energy in Your Life.

Physical Activity

Physical activity is defined as any bodily movement produced by skeletal muscles that requires energy expenditure, which encompasses moderate-vigorous exercise and lower-intensity walking [29]. A previous meta-analysis showed that physical activity reduced depression by a medium effect (SMD –0.50) and anxiety by a small effect (SMD –0.38) [30]. Doing physical activity is also recommended during the COVID-19 outbreak [31]. Imacoco Care provides knowledge and tips for leading an active life while people are asked to stay at home and being properly aware of the spread of the new coronavirus, in the following lessons: (1) Physical Activity for Mental Health—Introduction; 2) Let's Find a Physical Activity That Suits You—for Beginners; and (3) Let's Find a Physical Activity That Suits You—for Advanced Learners.

Sleep Education

Some studies have shown that sleep problems or insomnia were detected in around 35% of participants (ie, community residents or health care workers) during the COVID-19 pandemic [32,33]. CBT for insomnia (CBT-I) is one of the effective treatments for improving sleep problems, and it is recommended as the first-choice treatment for insomnia in the COVID-19 outbreak [34]. Imacoco Care provides information about how to regulate the sleep rhythm without using medication in the following lessons: (1) Psychological Advice for Those Whose Sleep Cycle Has Become Worse and (2) Life Tips to Try When You Can't Sleep.

Tips for Working From Home

Previous literature reviews have reported that the effects of working at home are inconsistent and might depend on various factors, such as demands of the home environment, the level of organizational support, and social connections external to work, among others [35]. This module provides some tips to make telecommuting (working from home) a little more comfortable, such as ergonomic tips when teleworking, keeping the time of the lunch break, getting fresh air, stretching and exercising, and keys for effective communication [36,37].

Coping With the Stress About COVID-19

This module provides stress management tips that can be practiced in daily life, including (1) prioritize personal hygiene (preventive measures), (2) stay connected with others, (3) keep a healthy lifestyle, (4) use psychological coping strategies (eg, mindfulness), (5) limit media consumption, and (6) prevent stigma and discrimination [36,38].

Intervention Group

Immediately after the baseline survey, participants in the intervention group were invited to access Imacoco Care and requested to view Imacoco Care within 1 month after the baseline survey. During this period, the participants received a reminder email twice (1 week after the baseline survey and 1 week before the 1-month follow-up survey).

Control Group

Participants in the control group were only asked to complete the baseline and follow-up surveys. They were not invited to the Imacoco Care website, but they were not restricted from accessing any other information or service as treatment as usual (TAU).

Outcome Measures

All outcomes were measured using a web-based self-report questionnaire at baseline, 1-month follow-up, and 3-month follow-up.

Primary Outcomes

Psychological Distress

The Kessler Psychological Distress Scale (K6) consists of 6 items assessing the frequency with which participants experience symptoms of psychological distress during the past 30 days [39,40]. The response options range from 0 (none of the time) to 4 (all the time). The internal reliability and validity found in previous studies are acceptable [39].

Fear of COVID-19 Infection

The Fear of COVID-19 Scale (FCV-19S) consists of 7 items assessing the fear of COVID-19 [41,42]. The response options range from 1 (strongly disagree) to 5 (strongly agree), with total scores ranging from 7 to 35. The higher the score, the greater the fear of COVID-19. The internal reliability and validity found in previous studies are acceptable [43].

Process Evaluation

Implementation Aspects of the Intervention Program, the Number of Views, and Progress in Learning About Imacoco Care

The Implementation Outcome Scale for Digital Mental Health (iOSDMH) for users was used to assess the implementation aspects of Imacoco Care [44]. The iOSDMH for users consists of 19 items assessing acceptability (3 items), appropriateness (4 items), feasibility (6 items), satisfaction (1 item), and harm (5 items). The response options range from 1 (disagree) to 4 (agree). The number of views was evaluated by using 1 item: "How many times did you view Imacoco Care during the past month?" There were 4 response options (never, once, twice, and 3/4 times or more). This scale was originally developed and has not yet been validated. To assess progress in learning about each of the 6 modules of Imacoco Care, participants in the intervention group were also asked, "Have you read and tried each content of Imacoco Care?" There were 3 response options (have not read, only read, and read and tried). This scale was also originally developed and has not yet been validated. These questions were asked of only participants in the intervention group in the 1-month follow-up survey.

Contamination of Information

To evaluate the degree of contamination of information, the control group participants were asked the following question in 1- and 3-month follow-up surveys: "During the past, have you ever viewed a website named Imacoco Care?" There were 4 response options (none, once, twice, and 3/4 times or more). This scale was originally developed and has not yet been validated.

Demographic Characteristics

Demographic data, such as age, gender, marital status (never married, married, divorced, or bereaved), education (high school or lower, some college, university, graduate school or higher), occupational status (manager, professional, clerical, production, sales, others), working style (working from home only, both working from home and at the office, working at the office only), and chronic disease (yes/no) were collected in the baseline survey.

Sample Size Calculation

The required sample size was calculated for psychological distress assessed by K6. A previous meta-analysis of an unguided eHealth intervention on improving workers' mental health in the workplace yielded an ES of 0.22 [45]. To detect a small ES (ie, 0.2) or more at an α error rate of .05 and a β error

rate of .10, the estimated sample size was 527 participants in each group. The statistical power was calculated using the G*Power 3.1 program [46,47].

Randomization

Participants who fulfilled the eligibility criteria were randomly allocated to 1 of 2 arms (intervention or control group). Stratified permuted-block randomization was conducted. The block sizes of this study were fixed to 2. Participants were stratified into 2 strata according to the K6 score (<4 or ≥5) in the baseline survey [48]. A stratified permuted block random table was generated by an independent biostatistician. Enrollment was conducted by a clinical research coordinator, and assignment was conducted by an independent research assistant. The stratified permuted-block random table was password-protected and blinded to the researcher. Only the research assistant could access it during the work of random allocation.

Statistical Methods

For the main analysis, a mixed model for repeated measures (MMRM) analysis of variance model was conducted using a group (intervention and control) × time (baseline, 1-month follow-up, and 3-month follow-up) interaction as an indicator of the intervention effect. Missing values at follow-up surveys were imputed applying the maximum likelihood estimation using the mixed procedure. An intention-to-treat (ITT) principle was applied.

The ES was estimated in 2 ways. First, we estimated a regression coefficient for a group (intervention vs control group) × time (baseline and 2 follow-ups) interaction using the mixed procedure, which was converted to an ES by dividing by a pooled SD at baseline and at follow-ups. Second, we calculated the Cohen d values [49] among participants who completed surveys at baseline and at each follow-up. The level of statistical significance for all analyses in this study was set at .05 (two-tailed), and 95% CIs were calculated. All statistical analyses were conducted using SPSS Statistics V.26.0 (IBM Corp).

Per Protocol Analysis

For the per protocol analysis, the same MMRM analysis was conducted using only participants who had viewed the Imacoco Care website at least once in the intervention group and all participants in the control group.

Results

Participant Recruitment

The participant flowchart is shown in Figure 1. Recruitment and the baseline survey were conducted in December 2020. Follow-up surveys in both groups were conducted 1 month (January 2021) and 3 months (March 2021) after the baseline survey. Participants who had fulfilled the inclusion criteria (ie, aged over 20 years and full-time employed) were recruited from monitors of an internet survey company (N=9484). Of those, 1367 (14.41%) were excluded according to the exclusion criteria. Of the 8117 eligible participants, 1200 (14.75%) who completed the baseline survey were selected on a first-come-first-served basis. Participants were randomly allocated to an intervention or a control group, with 600 (50%) participants in each.

Figure 1. Participant flow.
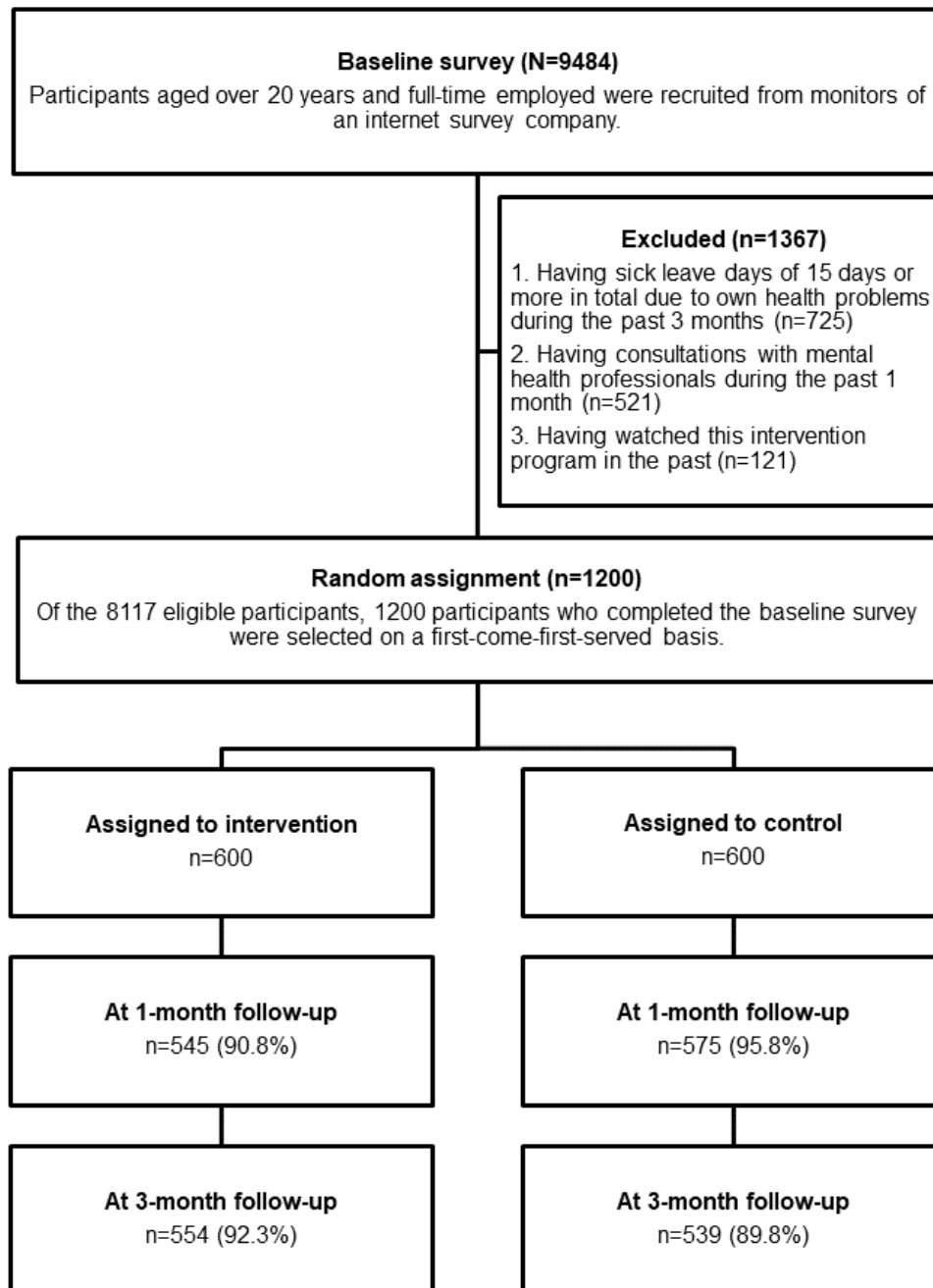
Baseline Characteristics

Demographic characteristics are presented in Table 2. In the whole sample, most participants were male, were married, received university or higher education, and did not report having chronic diseases. Only less than 30% were using telecommuting. Between the 2 groups, demographic characteristics of participants were similar. About 192 (32%) participants in each group had psychological distress (ie, scored 5 or more on K6).

Table 2. Baseline characteristics of participants in the intervention and control groups.

Variables	Intervention (N=600)	Control (N=600)
Age (years), mean (SD)	46.4 (10.3)	46.7 (10.0)
Gender, n (%)		
Male	441 (73.5)	442 (73.7)
Female	159 (26.5)	158 (26.3)
Marital status, n (%)		
Never married	190 (31.7)	190 (31.7)
Married	362 (60.3)	369 (61.5)
Divorced/bereaved	48 (4.0)	41 (6.8)
Education, n (%)		
High school	117 (19.5)	125 (20.8)
Some college	105 (17.5)	108 (18.0)
University	312 (52.0)	308 (51.3)
Graduate school	66 (11.0)	59 (9.8)
Occupation, n (%)		
Manager	122 (20.3)	135 (22.5)
Professional	186 (31.0)	164 (27.3)
Clerical	157 (26.2)	140 (23.3)
Production	57 (9.5)	66 (11.0)
Sales	61 (10.2)	64 (10.7)
Others	17 (2.8)	31 (5.2)
Working style, n (%)		
Working from home only	27 (4.5)	28 (4.7)
Both working from home and at the office	133 (22.2)	136 (22.7)
Working at the office only	440 (73.3)	436 (72.7)
Chronic disease, n (%)		
Yes	116 (19.3)	100 (16.7)
No	484 (80.7)	500 (83.3)
5 or more on K6 score	192 (32.0)	193 (32.2)

Effect of the Imacoco Care Program on Psychological Distress

The average K6 and FCV-19S scores in each survey are shown in Table 3. The K6 score increased at 1-month follow-up in both groups and decreased at 3-month follow-up only in the intervention group. The Imacoco Care intervention group showed a significant ES in reducing the K6 score at 3-month follow-up compared to the control group (Cohen d=–0.14, 95% CI –0.26 to –0.02). The FCV-19S score also increased at 1-month follow-up and decreased at 3-month follow-up in both groups. The Imacoco Care intervention group showed a significant adverse ES in increasing the FCV-19S score at 3-month follow-up compared to the control group (Cohen d=0.16, 95% CI 0.04-0.28).

Table 4 shows the estimated effects of the Imacoco Care intervention on K6 and FCV-19S scores on the basis of MMRM analyses. In ITT analyses, the Imacoco Care intervention showed a significant favorable effect on reducing the K6 score (P=.03) and an adverse effect on the FCV-19S score (P=.01) in the intervention group at 3-month follow-up. In the per protocol analyses, the Imacoco Care intervention also showed a significant favorable effect on reducing the K6 score (P=.03), while the adverse effect on the FCV-19S score was not significant (P=.06) in the intervention group at 3-month follow-up.

Table 3. Means (SDs) of outcome variables at baseline, 1-month follow-up, and 3-month follow-up in the intervention and control groups.

Outcome variables	Intervention		Control		ES[a]
	n (%)	Mean (SD)	n (%)	Mean (SD)	Cohen d (95% CI)
Psychological distress					
Baseline	600 (100)	3.8 (4.7)	600	3.7 (4.4)	N/A[b]
1-month follow-up	545 (90.8)	4.2 (5.0)	575	4.4 (5.0)	–0.07 (–0.19 to 0.04)
3-month follow-up	554 (92.3)	3.8 (4.9)	539	4.3 (5.2)	–0.14 (–0.26 to –0.02)
Fear about COVID-19					
Baseline	600 (100)	14.6 (5.6)	600	15.4 (5.8)	N/A
1-month follow-up	545 (90.8)	15.9 (5.8)	575	16.3 (5.9)	0.10 (–0.02 to 0.22)
3-month follow-up	554 (92.3)	14.8 (5.8)	539	14.8 (6.0)	0.16 (0.04-0.28)

[a]ES: effect size. Cohen d values were calculated among participants who answered each follow-up survey.

[b]N/A: not applicable.

Table 4. Effects of the Imacoco Care on outcomes.

Outcome variables	ITT[a]					Per protocol[b]				
	Effect (ES[c])	95% CI	SE	t (df)	P value	Effect (ES)	95% CI	SE	t (df)	P value
Psychological distress										
1 month	–0.30 (–0.08)	–0.76 to 0.17	0.24	–1.26 (1807.72)	.21	–0.39 (–0.10)	–0.98 to 0.20	0.30	–1.28 (1333.94)	.20
3 months	–0.53 (–0.13)	–1.01 to –0.06	0.24	–2.21 (1157.11)	.03	–0.67 (–0.17)	–1.30 to –0.05	0.32	–2.12 (797.91)	.03
Fear about COVID-19										
1 month	0.46 (0.10)	–0.10 to 1.03	0.29	1.60 (1847.74)	.11	0.60 (0.13)	–0.13 to 1.33	0.37	1.61 (1328.98)	.11
3 months	0.83 (0.16)	0.23-1.42	0.30	2.74 (1186.92)	.01	0.74 (0.15)	–0.03 to 1.52	0.39	1.89 (822.80)	.06

[a]ITT: intention to treat.

[b]Participants who saw the Imacoco Care website at least 1 time in the intervention group (n=235 [39.2%]) were included.

[c]ES: effect size, calculated by dividing the estimated effect by a pooled SD at baseline and at follow-ups.

Process Evaluation

Among participants (n=545, 90.8%) in the intervention group at the 1-month follow-up, 52 (9.5%) viewed the Imacoco Care website 4 times or more, 96 (17.6%) viewed it 2 or 3 times, 87 (16.0%) viewed it once, and 310 (56.9%) never viewed the Imacoco Care website.

Among participants (n=235, 39.2%) who viewed the Imacoco Care website at least once, 13 (5.5%) read all the content, 82 (34.9%) read most of the content, 73 (31.1%) read some of the content, 66 (28.1%) read a little of the content, and 1 (0.4%) read little of the content. In addition, 174 (74.0%) read or tried mindfulness techniques, 159 (67.7%) read or tried BA, 170 (72.3%) read or tried physical activity, 195 (83.0%) read or tried sleep education, 121 (51.4%) read or tried the tips for working from home, and 165 (70.2%) read or tried techniques for coping with stress about COVID-19.

Regarding the result of implementation outcomes assessed by the iOSDMH, around 105-174 (44.7%-74.0%) of the 235

participants answered "agree" or "relatively agree" on each item of acceptability (n=105-158, 44.7%-67.2%), appropriateness (n=115-164, 48.9%-69.8%), and feasibility (n=135-174, 57.4%-74.0%). The highest proportions of positive responses in each of the 3 aspects were "Advantages outweigh the disadvantages for keeping my mental health" (n=158, 67.2%) in acceptability, "Appropriate (from your perspective, it is the right thing to do)" (n=164, 69.8%) in appropriateness, and "Lower physical effort" (n=174, 74.0%) in feasibility, respectively. Regarding overall satisfaction with Imacoco Care, 135 (57.4%) answered "agree" or "relatively agree." Regarding harms, around 47 (20%) or fewer participants answered "agree" or "relatively agree" on each item, except for "Time-consuming" (n=63, 26.8%). The details of the number of participants for each item are shown in Multimedia Appendix 1.

Regarding the contamination of information, 9 (1.5%) of 600 participants in the control group had viewed the Imacoco Care website at least once at 1-month follow-up and 17 (2.83%) had viewed the Imacoco Care website at 3-month follow-up.

XSL•FO

RenderX

Discussion

Principal Findings

This large-scale RCT showed that the newly developed psychoeducational website during the COVID-19 pandemic, named Imacoco Care, significantly improved psychological distress at 3-month follow-up among workers in Japan, with a small ES. In concordance with previous studies conducted before the COVID-19 pandemic, a web-based psychoeducation approach including information about evidence-based psychological interventions may be effective for improving psychological distress (ie, depressive and anxiety symptoms) among workers in the current COVID-19 pandemic. The Imacoco Care website may promote mental health among workers, which mainly consisted of male, married, university graduates without chronic disease or current mental health problems, during the COVID-19 pandemic.

Comparison With Prior Work

To our knowledge, this study is the first to examine whether a psychoeducational website can improve psychological distress symptoms among workers during the COVID-19 pandemic. A significant intervention effect of the Imacoco Care intervention was found on psychological distress at 3-month follow-up, while the ES was small (Cohen d=–0.16) compared to previous studies of participants with elevated psychological distress or worry about COVID-19 (around 0.7) [14,15]. One of the possible reasons for the small ES in this RCT may be the lower intensity of the intervention. A previous RCT provided 5 modules with a few tasks to practice. Participants were encouraged to report on their progress in digital worksheets in the online platform [14]. The other pilot RCT provided 7 selected modules out of 16 possible modules, with a therapist's (ie, clinical psychologist) support. Each participant had a therapist who provided online support and feedback on the work with the modules and exercises and motivated the participant to continue to work with the treatment [15]. The Imacoco Care program only provided information via the website, and participants received a reminder email only twice. This may have led to a small ES; nevertheless, the Imacoco Care program still has merit because of its greater accessibility and lower cost. The other possible reason may be the difference in the baseline characteristics of the participants. The study excluded workers with current mental health problems (ie, having 15 sick leave days or more during the past 3 months or consultations with mental health professionals during the past month). This may have led to an underestimation of the intervention effect (floor effect).

Regarding fear about COVID-19, Imacoco Care showed a significant adverse effect (Cohen d=0.16) at 3-month follow-up. This slight increase in the fear about COVID-19 may be due to increased awareness of stress caused by learning from the Imacoco Care program [13,50]. In contrast, the result from per protocol analysis that included only participants who had read the Imacoco Care website at least 1 time showed a nonsignificant adverse effect on the fear of COVID-19 at 3-month follow-up (Cohen d=0.15). An adverse effect on the fear about COVID-19 may have been more common among participants who never saw the Imacoco Care website during

the intervention period. Both distributing the information and encouraging people to access and read the content may be necessary to prevent adverse effects of psychoeducational intervention. Users should be informed about the possible adverse effects of seeing the Imacoco Care website in advance to decide whether to take the program, balancing merits and demerits.

Process evaluation of the Imacoco Care program showed that the proportion of those who viewed the Imacoco Care website at least once was 235 (39.2%) of 600 participants. Of those, only 95 (40.4%) read all or most of the content on the Imacoco Care website. The low rates of reading the content may weaken the intervention effect. Regarding the results from per protocol analyses, the Imacoco Care program showed a greater effect for psychological distress (Cohen d=–0.17) among workers who had viewed the Imacoco Care program at least once. Increasing readers may be effective in improving the effect of Imacoco Care. From the implementation aspects, Imacoco Care showed moderate rates (around 105-174 [44.7%-74.0%]) of acceptability, appropriateness, and feasibility among participants who viewed the Imacoco Care website at least once. The features of the Imacoco Care program (explaining evidence-based psychological intervention in an easy-to-understand way by using plain language, illustrations, video, and audio narration) may contribute to these implementation outcomes. Further study is needed to develop an implementation strategy for increasing adoption of the Imacoco Care program among workers and examine its effect on implementation outcomes.

Limitations

Several limitations of this study should be considered. First, all participants were recruited from registered members of a web survey company in Japan. Most of participants were male, married, and university graduates without chronic disease and psychological distress at the baseline survey. Therefore, generalization of the findings to the general working population may be limited. Second, website access logs by users were not collected. The actual proportion of access to each content area of Imacoco Care could not be analyzed. The effectiveness of each content area could not be examined, and the results from process evaluation may be biased. Third, all outcomes were measured by self-report, which may be affected by the perceptions of the participants or by situational factors at work. A further RCT and implementation study should be conducted to examine whether psychoeducational intervention via an online program is effective among workers with diverse characteristics, particularly in terms of education.

Conclusion

An RCT was used to evaluate a new website-based psychoeducational program, named Imacoco Care, which was developed to address mental health issues during the COVID-19 pandemic. Results showed significantly reduced psychological distress at 3-month follow-up among workers in Japan, with a small ES. Thus, a web-based psychoeducation approach may be an effective intervention for reducing psychological distress in the working population. To enhance outcomes from this type of intervention and reduce any adverse effects, it may be important to not only distribute information about the program

XSL•FO

RenderX

but also encourage workers to access and actively engage with   the content of the program.

Acknowledgments

Authors' Contributions

KI, NS, YS, KW, AS, YM, DN, and NK contributed to conception and design of the study; KI, NS, and NK contributed to acquisition and analysis of data; and KI, NS, and NK contributed to drafting the manuscript and figures. All authors have read and approved the final paper.

Conflicts of Interest

NK reports grants from Fujitsu Ltd and TAK Ltd and personal fees from the Occupational Health Foundation, the Japan Dental Association, Sekisui Chemicals, the Junpukai Health Care Center, and the Osaka Chamber of Commerce and Industry, outside the submitted work.

Multimedia Appendix 1
Supplemental table.
[DOCX File , 19 KB - formative_v6i3e33883_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 350 KB - formative_v6i3e33883_app2.pdf]

References

1. Sun P, Wang M, Song T, Wu Y, Luo J, Chen L, et al. The psychological impact of COVID-19 pandemic on health care workers: a systematic review and meta-analysis. Front Psychol 2021;12:626547 [FREE Full text] [doi: 10.3389/fpsyg.2021.626547] [Medline: 34305703]
2. Saragih ID, Tonapa SI, Saragih IS, Advani S, Batubara SO, Suarilah I, et al. Global prevalence of mental health problems among healthcare workers during the Covid-19 pandemic: a systematic review and meta-analysis. Int J Nurs Stud 2021 Sep;121:104002. [doi: 10.1016/j.ijnurstu.2021.104002] [Medline: 34271460]
3. Olaya B, Pérez-Moreno M, Bueno-Notivol J, Gracia-García P, Lasheras I, Santabárbara J. Prevalence of depression among healthcare workers during the COVID-19 outbreak: a systematic review and meta-analysis. J Clin Med 2021 Jul 30;10(15):3406 [FREE Full text] [doi: 10.3390/jcm10153406] [Medline: 34362188]
4. Raoofi S, Pashazadeh Kan F, Rafiei S, Khani S, Hosseinifard H, Tajik F, et al. Anxiety during the COVID-19 pandemic in hospital staff: systematic review plus meta-analysis. BMJ Support Palliat Care 2021 Jul 26:1-9. [doi: 10.1136/bmjspcare-2021-003125] [Medline: 34312187]
5. Sahebi A, Abdi K, Moayedi S, Torres M, Golitaleb M. The prevalence of insomnia among health care workers amid the COVID-19 pandemic: an umbrella review of meta-analyses. J Psychosom Res 2021 Oct;149:110597 [FREE Full text] [doi: 10.1016/j.jpsychores.2021.110597] [Medline: 34388380]
6. Ruiz-Frutos C, Ortega-Moreno M, Allande-Cussó R, Domínguez-Salas S, Dias A, Gómez-Salgado J. Health-related factors of psychological distress during the COVID-19 pandemic among non-health workers in Spain. Saf Sci 2021 Jan;133:104996 [FREE Full text] [doi: 10.1016/j.ssci.2020.104996] [Medline: 32952305]
7. Smith PM, Oudyk J, Potter G, Mustard C. Labour market attachment, workplace infection control procedures and mental health: a cross-sectional survey of Canadian non-healthcare workers during the COVID-19 pandemic. Ann Work Expo Health 2021 Apr 22;65(3):266-276 [FREE Full text] [doi: 10.1093/annweh/wxaa119] [Medline: 33313670]
8. Wong EL, Ho K, Wong SY, Cheung AW, Yau PS, Dong D, et al. Views on workplace policies and its impact on health-related quality of life during coronavirus disease (COVID-19) pandemic: cross-sectional survey of employees. Int J Health Policy Manag 2022 Mar 01;11(3):344-353. [doi: 10.34172/ijhpm.2020.127] [Medline: 32772010]
9. Carbone SR. Flattening the curve of mental ill-health: the importance of primary prevention in managing the mental health impacts of COVID-19. Ment Health Prev 2020 Sep;19:200185 [FREE Full text] [doi: 10.1016/j.mhp.2020.200185] [Medline: 32566473]
10. Zaçe D, Hoxhaj I, Orfino A, Viteritti AM, Janiri L, Di Pietro ML. Interventions to address mental health issues in healthcare workers during infectious disease outbreaks: a systematic review. J Psychiatr Res 2021 Apr;136:319-333 [FREE Full text] [doi: 10.1016/j.jpsychires.2021.02.019] [Medline: 33636688]

XSL•FO
RenderX

11. Rigabert A, Motrico E, Moreno-Peral P, Resurrección DM, Conejo-Cerón S, Cuijpers P, et al. Effectiveness of online psychological and psychoeducational interventions to prevent depression: systematic review and meta-analysis of randomized controlled trials. Clin Psychol Rev 2020 Dec;82:101931. [doi: 10.1016/j.cpr.2020.101931] [Medline: 33137611]

12. Christensen H, Griffiths KM, Jorm AF. Delivering interventions for depression by using the internet: randomised controlled trial. BMJ 2004 Jan 31;328(7434):265 [FREE Full text] [doi: 10.1136/bmj.37945.566632.EE] [Medline: 14742346]

13. Imamura K, Kawakami N, Tsuno K, Tsuchiya M, Shimada K, Namba K. Effects of web-based stress and depression literacy intervention on improving symptoms and knowledge of depression among workers: a randomized controlled trial. J Affect Disord 2016 Oct;203:30-37. [doi: 10.1016/j.jad.2016.05.045] [Medline: 27280960]

14. Wahlund T, Mataix-Cols D, Olofsdotter Lauri K, de Schipper E, Ljótsson B, Aspvall K, et al. Brief online cognitive behavioural intervention for dysfunctional worry related to the COVID-19 pandemic: a randomised controlled trial. Psychother Psychosom 2021 Nov 19;90(3):191-199 [FREE Full text] [doi: 10.1159/000512843] [Medline: 33212440]

15. Aminoff V, Sellén M, Sörliden E, Ludvigsson M, Berg M, Andersson G. Internet-based cognitive behavioral therapy for psychological distress associated with the COVID-19 pandemic: a pilot randomized controlled trial. Front Psychol 2021;12:684540 [FREE Full text] [doi: 10.3389/fpsyg.2021.684540] [Medline: 34194374]

16. Pheh K, Tan H, Tan C. Effects of an ultra-brief online mindfulness-based intervention on mental health during the coronavirus disease (COVID-19) outbreak in Malaysia: a randomized controlled trial. Makara Hubs-Asia 2020 Dec 31;24(2):118. [doi: 10.7454/hubs.asia.2140920]

17. Fiol-DeRoque MA, Serrano-Ripoll MJ, Jiménez R, Zamanillo-Campos R, Yáñez-Juan AM, Bennasar-Veny M, et al. A mobile phone-based intervention to reduce mental health problems in health care workers during the COVID-19 pandemic (PsyCovidApp): randomized controlled trial. JMIR Mhealth Uhealth 2021 May 18;9(5):e27039 [FREE Full text] [doi: 10.2196/27039] [Medline: 33909587]

18. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]

19. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

20. Rakuten Insight. URL: https://insight.rakuten.co.jp/en/ [accessed 2022-03-03]

21. Imacoco Care. URL: https://imacococare.net/en/ [accessed 2022-03-03]

22. Xiao Y, Becerik-Gerber B, Lucas G, Roll SC. Impacts of working from home during COVID-19 pandemic on physical and mental well-being of office workstation users. J Occup Environ Med 2021 Mar 01;63(3):181-190 [FREE Full text] [doi: 10.1097/JOM.0000000000002097] [Medline: 33234875]

23. Toniolo-Barrios M, Pitt L. Mindfulness and the challenges of working from home in times of crisis. Bus Horiz 2021;64(2):189-197 [FREE Full text] [doi: 10.1016/j.bushor.2020.09.004] [Medline: 33041346]

24. Kabat-Zinn J. Mindfulness-based interventions in context: past, present, and future. Clin Psychol Sci Pract 2003;10(2):144-156 [FREE Full text] [doi: 10.1093/clipsy.bpg016] [Medline: 000182986300002]

25. Goyal M, Singh S, Sibinga EMS, Gould NF, Rowland-Seymour A, Sharma R, et al. Meditation programs for psychological stress and well-being: a systematic review and meta-analysis. JAMA Intern Med 2014 Mar;174(3):357-368 [FREE Full text] [doi: 10.1001/jamainternmed.2013.13018] [Medline: 24395196]

26. Neff KD, Germer CK. A pilot study and randomized controlled trial of the mindful self-compassion program. J Clin Psychol 2013 Jan;69(1):28-44. [doi: 10.1002/jclp.21923] [Medline: 23070875]

27. Martell C, Dimidjian S, Herman-Dunn R. Behavioral Activation for Depression: A Clinician's Guide. New York, NY: Guilford Press; 2013.

28. Cuijpers P, Muñoz RF, Clarke GN, Lewinsohn PM. Psychoeducational treatment and prevention of depression: the "Coping with Depression" course thirty years later. Clin Psychol Rev 2009 Jul;29(5):449-458. [doi: 10.1016/j.cpr.2009.04.005] [Medline: 19450912]

29. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep 1985;100(2):126-131 [FREE Full text] [Medline: 3920711]

30. Rebar AL, Stanton R, Geard D, Short C, Duncan MJ, Vandelanotte C. A meta-meta-analysis of the effect of physical activity on depression and anxiety in non-clinical adult populations. Health Psychol Rev 2015;9(3):366-378. [doi: 10.1080/17437199.2015.1022901] [Medline: 25739893]

31. Jurak G, Morrison SA, Leskošek B, Kovač M, Hadžić V, Vodičar J, et al. Physical activity recommendations during the coronavirus disease-2019 virus outbreak. J Sport Health Sci 2020 Jul;9(4):325-327 [FREE Full text] [doi: 10.1016/j.jshs.2020.05.003] [Medline: 32426171]

32. Voitsidis P, Gliatas I, Bairachtari V, Papadopoulou K, Papageorgiou G, Parlapani E, et al. Insomnia during the COVID-19 pandemic in a Greek population. Psychiatry Res 2020 May 12;289:113076. [doi: 10.1016/j.psychres.2020.113076] [Medline: 33242799]

33. Pappa S, Ntella V, Giannakas T, Giannakoulis VG, Papoutsi E, Katsaounou P. Prevalence of depression, anxiety, and insomnia among healthcare workers during the COVID-19 pandemic: a systematic review and meta-analysis. Brain Behav Immun 2020 Aug;88:901-907 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.026] [Medline: 32437915]

34. Altena E, Baglioni C, Espie CA, Ellis J, Gavriloff D, Holzinger B, et al. Dealing with sleep problems during home confinement due to the COVID-19 outbreak: practical recommendations from a task force of the European CBT-I Academy. J Sleep Res 2020 Aug 04;29(4):e13052 [FREE Full text] [doi: 10.1111/jsr.13052] [Medline: 32246787]

35. Oakman J, Kinsman N, Stuckey R, Graham M, Weale V. A rapid review of mental and physical health effects of working at home: how do we optimise health? BMC Public Health 2020 Nov 30;20(1):1825 [FREE Full text] [doi: 10.1186/s12889-020-09875-z] [Medline: 33256652]

36. 34. American Psychiatric Association. Working Remotely During COVID-19: Your Mental Health and Well-Being. URL: https://www.workplacementalhealth.org/employer-resources/working-remotely-during-covid-19 [accessed 2022-03-03]

37. 35. International Labour Organization. Keys for Effective Teleworking during the COVID-19 Pandemic. URL: https://www.ilo.org/global/about-the-ilo/newsroom/news/WCMS_739879/lang--en/index.htm [accessed 2022-03-03]

38. Giorgi G, Lecca LI, Alessio F, Finstad GL, Bondanini G, Lulli LG, et al. COVID-19-related mental health effects in the workplace: a narrative review. Int J Environ Res Public Health 2020 Oct 27;17(21):7857 [FREE Full text] [doi: 10.3390/ijerph17217857] [Medline: 33120930]

39. Furukawa TA, Kawakami N, Saitoh M, Ono Y, Nakane Y, Nakamura Y, et al. The performance of the Japanese version of the K6 and K10 in the World Mental Health Survey Japan. Int J Methods Psychiatr Res 2008;17(3):152-158 [FREE Full text] [doi: 10.1002/mpr.257] [Medline: 18763695]

40. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SLT, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002 Aug;32(6):959-976. [doi: 10.1017/s0033291702006074] [Medline: 12214795]

41. Ahorsu DK, Lin C, Imani V, Saffari M, Griffiths MD, Pakpour AH. The Fear of COVID-19 Scale: development and initial validation. Int J Ment Health Addict 2020 Mar 27:1-9 [FREE Full text] [doi: 10.1007/s11469-020-00270-8] [Medline: 32226353]

42. Soraci P, Ferrari A, Abbiati FA, Del Fante E, De Pace R, Urso A, et al. Validation and psychometric evaluation of the Italian version of the Fear of COVID-19 Scale. Int J Ment Health Addict 2020 May 04:1-10 [FREE Full text] [doi: 10.1007/s11469-020-00277-1] [Medline: 32372892]

43. Wakashima K, Asai K, Kobayashi D, Koiwa K, Kamoshida S, Sakuraba M. The Japanese version of the Fear of COVID-19 scale: reliability, validity, and relation to coping behavior. PLoS One 2020;15(11):e0241958 [FREE Full text] [doi: 10.1371/journal.pone.0241958] [Medline: 33152038]

44. Sasaki N, Obikane E, Vedanthan R, Imamura K, Cuijpers P, Shimazu T, et al. Implementation Outcome Scales for Digital Mental Health (iOSDMH): scale development and cross-sectional study. JMIR Form Res 2021 Nov 23;5(11):e24332 [FREE Full text] [doi: 10.2196/24332] [Medline: 34817391]

45. Stratton E, Lampit A, Choi I, Calvo RA, Harvey SB, Glozier N. Effectiveness of eHealth interventions for reducing mental health conditions in employees: a systematic review and meta-analysis. PLoS One 2017;12(12):e0189904 [FREE Full text] [doi: 10.1371/journal.pone.0189904] [Medline: 29267334]

46. Faul F, Erdfelder E, Buchner A, Lang A. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/brm.41.4.1149]

47. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]

48. Sakurai K, Nishi A, Kondo K, Yanagida K, Kawakami N. Screening performance of K6/K10 and other screening instruments for mood and anxiety disorders in Japan. Psychiatry Clin Neurosci 2011 Aug;65(5):434-441 [FREE Full text] [doi: 10.1111/j.1440-1819.2011.02236.x] [Medline: 21851452]

49. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]

50. Shimazu A, Umanodan R, Schaufeli WB. Effects of a brief worksite stress management program on coping skills, psychological distress and physical complaints: a controlled trial. Int Arch Occup Environ Health 2006 Oct;80(1):60-69. [doi: 10.1007/s00420-006-0104-9] [Medline: 16633813]

Abbreviations

BA: behavioral activation
CBT: cognitive behavioral therapy
ES: effect size
FCV-19S: Fear of COVID-19 Scale
iOSDMH: Implementation Outcome Scales for Digital Mental Health
ITT: intention to treat
K6: Kessler Psychological Distress Scale
MMRM: mixed model for repeated measures
RCT: randomized controlled trial
SMD: standardized mean difference

XSL•FO
RenderX

Barriers and Facilitators to HIV Testing Among Adolescents and Young Adults in Washington, District of Columbia: Formative Research to Inform the Development of an mHealth Intervention

Brittany Wilbourn[1], MPH; Tyriesa Howard-Howell[2], LMSW, PhD; Amanda Castel[1], MPH, MD; Lawrence D'Angelo[3], MPH, MD; Constance Trexler[3], RN, CPN, MSHS; Rashida Carr[3], MHS; Daniel Greenberg[4], BA

[1]Department of Epidemiology, Milken Institute School of Public Health, The George Washington University, Washington, DC, United States
[2]Washington University in St. Louis Brown School, St, Louis, MO, United States
[3]Children's National Hospital, Washington, DC, United States
[4]Media Rez, Washington, DC, United States

Corresponding Author:
Brittany Wilbourn, MPH
Department of Epidemiology
Milken Institute School of Public Health
The George Washington University
950 New Hampshire Ave
Suite 500
Washington, DC, 20052
United States
Phone: 1 2029944392
Email: blewis1@gwu.edu

Abstract

Background: Adolescents and young adults (AYA) in the United States, and in Washington, District of Columbia (DC), specifically, are disproportionately affected by HIV. Both the national Ending the HIV Epidemic initiative and DC-specific plans emphasize HIV testing, and innovative strategies to encourage testing among AYA are needed.

Objective: The purpose of this study is to identify sexual behaviors, HIV knowledge, HIV perceptions (eg, susceptibility and severity), and perceived barriers and facilitators to HIV testing among AYA at risk for HIV in DC.

Methods: This study was part of a larger study to determine the acceptability of using a life-and-dating simulation game to increase HIV testing among AYA. Focus groups and surveys stratified by self-reported sexual orientation were conducted among, and administered to, AYA aged 13-24 years in DC. HIV knowledge was explored during focus groups and measured using an adapted version of the Brief HIV Knowledge Questionnaire. Survey data were summarized using descriptive statistics and compared by self-reported sexual orientation. Transcripts were thematically analyzed.

Results: Of the 46 AYA who participated in the focus groups, 30 (65%) identified as heterosexual and 16 (35%) as lesbian, gay, bisexual, transgender, or queer. A higher proportion of lesbian, gay, bisexual, transgender, or queer AYA reported sexual activity (12/16, 75%, vs 18/30, 60%), condomless sex (11/12, 92%, vs 15/18, 83%), and HIV testing (13/16, 81%, vs 17/29, 58%) than heterosexual AYA. HIV prevention ("condoms" and "...PrEP") and transmission ("exchange of fluids") knowledge was high, and most (34/44, 77%) of the AYA perceived HIV testing as beneficial. However, the AYA also demonstrated some misinformation concerning HIV: an average of 67% (31/46; SD 0.474) of the participants believed that an HIV test could deliver accurate results 1 week after a potential exposure and an average of 72% (33/46; SD 0.455) believed that an HIV vaccine exists. The AYA also identified individual ("...people...are scared"), interpersonal ("it's an awkward conversation"), and structural ("...people don't...know where they can go") barriers to testing. Most of the AYA indicated that they were very likely to use the demonstrated game prototype to help with getting tested for HIV (median 3.0, IQR 2.0-3.0, using a scale ranging from 0 to 3, with 3 indicating high likelihood) and strongly agreed that the game was interesting (median 5.0, IQR 5.0-5.0), fun (median 5.0, IQR 4.0-5.0), and easy to learn (median 5.0, IQR 5.0-5.0, using a scale ranging from 1 to 5, with 5 indicating strong agreement).

Conclusions: These results suggest a need for multilevel HIV testing interventions and informed the development of a mobile health intervention aiming to increase HIV knowledge and risk perception among AYA, while reducing barriers to testing at the individual and structural levels, supporting efforts to end the domestic HIV epidemic.

Introduction

HIV Among Adolescents and Young Adults in the United States

Adolescents and young adults (AYA) aged 13-24 years are disproportionately affected by HIV in the United States. Among the estimated HIV transmissions in 2016, AYA had the highest transmission rate of all age groups [1]. In 2019, AYA accounted for 21% of the new HIV infections, and young men who have sex with men (YMSM) and transgender women who identified as Black or Latinx were severely affected [2]. Approximately 50% of the AYA living with HIV in the United States are unaware of their infection and are more likely to be unaware of their infection than any other age group [2,3]. Nevertheless, once AYA are diagnosed, they often reduce their sexual risk behaviors, which limits the risk of infecting others [4-6]. In addition, AYA who are diagnosed with HIV early in the course of their infection can access antiretroviral therapy to slow disease progression and reduce mortality [4].

HIV Among AYA in Washington, District of Columbia

Just as HIV infections and diagnoses in the United States are not evenly distributed across different ages, races and ethnicities, genders, and modes of transmission, they are more concentrated in certain geographic locations than in others. The southern United States, which includes Washington, District of Columbia (DC), accounted for 53% of the new diagnoses in 2019 despite being home to just 38% of the population [7]. In DC specifically—a metropolitan area that also includes parts of Maryland, Virginia, and West Virginia—AYA are disproportionately affected by HIV, with those aged 13-24 years accounting for 21% of the new diagnoses in 2019 [8]. Furthermore, Black AYA and YMSM in DC accounted for 75% and 60%, respectively, of the new HIV diagnoses among AYA in 2019 [9]. The recently launched *Ending the HIV Epidemic in the U.S.* (EHE) initiative aims to end the domestic HIV epidemic by 2030, and the first strategy includes ensuring that HIV testing is widely available to diagnose infections as early as possible [10]. DC is among several localities across the nation considered geographic HIV hotspots that will be the focus of initial EHE efforts [10]. The DC-specific plan to end the HIV epidemic also includes emphasis on HIV testing and timely diagnosis [11], and DC's Youth Sexual Health Plan [12] has the explicit goal of reducing the unintended consequences of condomless sex among AYA (eg, HIV infection) by increasing use of sexual and reproductive health services, including HIV testing.

Lack of HIV Testing Among AYA

Despite engaging in high-risk behaviors [13-19] and the existence of national guidelines that support routine HIV testing among persons aged 13-64 years [19,20], AYA do not regularly seek nor are they routinely offered testing for HIV. Barriers to HIV testing among AYA include privacy concerns, parental involvement, inconvenient clinic times, providers not assessing sexual behavior, cost, low access, and low health literacy [21-25]. Furthermore, AYA may have limited knowledge of HIV [26-29] and underestimate their risk for infection [21,25,30]. In DC, differences in perceived access to HIV education exist among AYA based on sexual orientation. According to the 2019 Youth Risk Behavior Survey (YRBS), heterosexual middle school students in DC were more likely to report either not being taught or not being sure that they were taught about HIV/AIDS in school than lesbian, gay, and bisexual middle school students [31]. YRBS data also suggest that high-risk behaviors among youth in DC differ based on sexual orientation. Among middle school and high school students who reported sexual activity on the 2019 YRBS, lesbian, gay, and bisexual students were more likely to report having condomless sex during their last sexual encounter than heterosexual students [31,32].

Potential of Mobile Health Interventions and Study Objectives

Innovative strategies to overcome these barriers and increase HIV testing among AYA include social media campaigns [33,34] and mobile health (mHealth) interventions [35,36], both of which have been endorsed by the Centers for Disease Control and Prevention (CDC) to reduce HIV risk and improve sexual health among adolescents [37]. New mHealth interventions for HIV treatment and prevention have been developed specifically for AYA in urban areas [35,36,38-42]. These interventions transcend traditional behavior modification approaches to HIV prevention and incorporate adolescent-friendly and -informed models that support the design and development of mHealth interventions. The purpose of this study is to identify sexual behaviors, HIV knowledge, HIV perceptions (eg, susceptibility and severity), and perceived barriers and facilitators to HIV testing among AYA at risk for HIV and to identify potential differences based on self-reported sexual orientation. The results from this study were used to inform the development of an mHealth intervention aiming to increase HIV testing among AYA living in DC.

Methods

Data Collection

The study was conducted as part of a larger study to determine the acceptability of using a life-and-dating simulation game to increase HIV testing among AYA; the methods have been described elsewhere [43]. The game will allow AYA to play characters that are similar to, or different from, their gender identity. The character can then spend their leisure time meeting new people and engaging in relationship scenarios—including implied sexual activity—tailored to both same-sex and opposite-sex couples. On the basis of in-game sexual activity, AYA will be shown their character's risk of acquiring HIV and be allowed to locate nearby AYA-friendly testing locations.

Guided by the Health Belief Model [44]—a widely used framework to explain and predict individual-level changes to health-related behavior [44] and to assess HIV testing uptake specifically [45-48]—we hypothesized that the game would increase perceived susceptibility to HIV infection, perceived severity of HIV infection, perceived benefits of HIV testing, motivation to test, and self-efficacy to test while decreasing perceived barriers to testing among the study participants (Figure 1).

Figure 1. Conceptual framework for assessing HIV testing uptake among adolescents and young adults. PrEP: pre-exposure prophylaxis; SES: socioeconomic status.
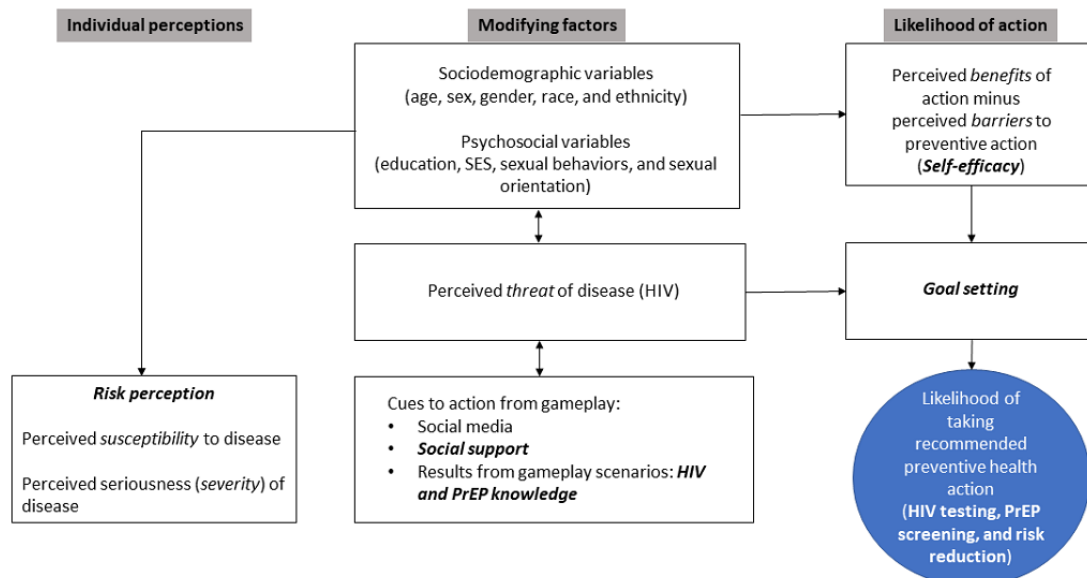
From February 2017 to December 2017, HIV-negative or status-unknown AYA aged 13-24 years who were living in the DC metro area were recruited from an adolescent health center during routine visits, 2 community-based organizations, and a youth advisory council [43]. Recruitment was primarily facilitated by CT, a female researcher. Of the 52 AYA approached for the study, 6 (12%) declined to participate; the reasons for refusal included lack of time to undergo the eligibility screening process or not being available on the date of the focus group. To assess the baseline perceptions of AYA regarding HIV susceptibility and severity as well as potential benefits of, and barriers to, testing, we conducted a series of focus group discussions that lasted 50-120 minutes and were held at the adolescent health center, a local community-based organization, and a DC government building where the youth advisory council met. Focus group discussions, stratified by sexual orientation, were facilitated by ADC (a female researcher) using a semistructured focus group guide and covered the following topics: HIV knowledge, perceptions and personal experiences regarding HIV testing, and perceived barriers and facilitators to HIV testing. The focus group guide can be found in Multimedia Appendix 1.

Next, DG (a male researcher) demonstrated the game to participants and asked questions related to game acceptability. The game showed two dating scenarios: one at a house party and one at a skateboard park. DG controlled a character that had the freedom to go wherever they wanted in the 3D environment and talk to other AYA in the game by selecting from a menu of dialogue options. These conversations, combined with the player's actions, resulted in the player's character becoming friends, rivals, or dating partners with the people they met. Participants played the game as a group by voting on the actions and dialogue options the player character should choose in navigating dating scenarios. When the player managed to negotiate sex with another character in the game, the participants chose the kinds of sex acts the characters would engage in and whether a condom would be used. The sex acts were not depicted graphically but rather implied by fireworks, and the screen abruptly transitioned to a depiction of the risk of the sex acts chosen based on the CDC Risk Estimator Tool. The player then had the option of seeing a map of nearby HIV testing providers, with the ability to select low-cost and free options. Participants could replay the game, trying other options and seeing other outcomes.

Focus groups continued until data saturation—the point at which new data repeat data already collected [49]—was met, which occurred after 7 focus groups. Before the start of each focus group, participants completed a survey that collected information on their sociodemographic characteristics, HIV knowledge and HIV testing patterns, sexual behaviors and perceptions regarding susceptibility to HIV, self-efficacy to get tested for HIV, and barriers and benefits to HIV testing. As in the case of other research assessing HIV knowledge using a limited number of questions [50,51], HIV knowledge in this study was measured using an adapted version of the Brief HIV Knowledge Questionnaire that included 11 of the 18 questions [52]. After the game demonstration, the AYA completed a brief survey to assess game acceptability. All AYA participants were remunerated with gift cards and transportation vouchers for their participation in each focus group.

Data Analysis

All focus groups were audio recorded and professionally transcribed, and transcripts were imported into NVivo 11 (QSR

International) for data analysis. Initial codes were deductively derived: BW created a codebook containing a priori codes based on the topics included in the semistructured focus group guide, and 2 coders (BW and THH) conducted independent thematic coding to link the common themes and ideas discussed among study participants in each respective focus group [53]. Next, the coders reconvened to reach intercoder agreement through reconciliation of discrepant coding. Specific themes across focus group data were developed based on patterns of shared or similar meaning identified through codes ascribed to each respondent's feedback. Finally, final codes across focus groups and coders were compared to generate overall themes that supported the purpose of the study. In addition to focus groups, data collected from surveys were summarized using descriptive statistics in SAS 9.3 (SAS Institute Inc) by BW. Survey responses related to demographics, sexual behavior, and risk perception were stratified by self-reported sexual orientation; frequencies and percentages were used to describe categorical variables, whereas mean and range or mean and SD were used to describe continuous variables. To describe acceptability of the game prototype, survey responses were stratified by self-reported sexual orientation and median values were reported for each survey item.

Ethics Approval

Study materials were reviewed and approved by the George Washington University Institutional Review Board (IRB) (number 051602) and the Children's National Hospital IRB.

Results

AYA Demographics, Sexual Risk Behaviors, and Perceptions

Of the 46 AYA who participated in the focus groups, 30 (65%) identified as heterosexual and 16 (35%) identified as lesbian, gay, bisexual, transgender, or queer (LGBTQ; Table 1). The mean age of participants was 17.6 (SD 2.110; range 13-24) years, 85% (39/46) were Black, and 63% (29/46) were female. Most of the participants lived in DC (44/46, 96%), and 67% (30/45) had not yet graduated from high school. Higher proportions of the LGBTQ participants reported engaging in sexual activity (12/16, 75%, vs 18/30, 60%), having a history of condomless sex (11/12, 92% vs 15/18, 83%), and previous testing for HIV (13/16, 81%, vs 17/29, 59%). Most (41/44, 93%) of the participants perceived HIV infection to be severe or very severe (Table 1).

Regarding their perceived risk of acquiring HIV, on a scale of 0%-100%, the reported likelihood of HIV infection among heterosexual AYA (mean 13%, SD 28%) was similar to that of LGBTQ AYA (mean 14%, SD 27%). Most (34/44, 77%) of the participants perceived HIV testing to be beneficial or very beneficial (Table 1). Furthermore, most of the participants perceived finding a testing location (38/46, 83%), paying for testing (23/45, 51%), and disclosing their sexual orientation during testing (27/45, 60%) to be easy or very easy.

Table 1. Demographics, HIV testing, and risk behaviors of adolescents and young adults and their peers (N=46).

Demographics and sexual behaviors	Heterosexual (n=30)[a]	LGBTQ[b] (n=16)[a]	Total (N=46)[a]
Age (years), mean (SD; range)	17.1 (1.776; 13-23)	18.6 (2.391; 15-24)	17.6 (2.110; 13-24)
Gender, n (%)			
Male	10 (33)	6 (38)	16 (35)
Female	20 (67)	9 (56)	29 (63)
Transgender	0 (0)	1 (6)	1 (2)
Race and ethnicity, n (%)			
Hispanic	2 (7)	1 (6)	3 (6)
Non-Hispanic Black	26 (87)	13 (81)	39 (85)
Non-Hispanic White	1 (3)	0 (0)	1 (2)
Other	1 (3)	2 (13)	3 (6)
State of residence, n (%)			
Washington, District of Columbia	28 (93)	16 (100)	44 (96)
Maryland	2 (7)	0 (0)	2 (4)
Highest level of education, n (%)			
Grade 8-11	23 (77)	7[c] (47)	30[d] (67)
High school graduate	6 (20)	5[c] (33)	11[d] (24)
At least some college	1 (3)	3[c] (20)	4[d] (9)
Ever tested for HIV, n (%)	17[e] (58)	13 (81)	30[d] (67)
Location of most recent test: physician's office, n (%)	4[f] (23)	6[g] (46)	10[c] (33)
Reason for testing: offered a free test, n (%)	8[f] (47)	6[g] (46)	14[h] (47)
Reason never tested: not sexually active, n (%)	5[i] (50)	0[j] (0)	5[g] (38)
Ever sexually active, n (%)	18 (60)	12 (75)	30 (65)
Condomless sex ever, n (%)	15[k] (83)	11[l] (92)	26[h] (87)
Percentage of time spent... (VAS[m]), mean (SD)			
Engaging in risky behavior	8 (14)	12 (22)	10 (77)
Using condoms consistently	63 (38)	40 (38)	54 (39)
Risk perception (perceived benefit of HIV testing: beneficial or very beneficial), n (%)	23 (76)	11[n] (79)	34[o] (77)
Perceived ease of...[p], n (%)			
Finding a testing location	27 (90)	11 (69)	38 (83)
Getting parental permission for testing	13 (43)	9 (56)	22 (48)
Paying for testing	18 (60)	5[c] (33)	23[d] (51)
Disclosing sexual orientation to health care provider or tester	20[q] (69)	7 (44)	27[d] (60)
Likelihood of infection (VAS), mean (SD)	13 (28)	14 (27)	13 (28)
Perceived severity of HIV infection: severe or very severe, n (%)	29 (97)	12[n] (86)	41[o] (93)
Percentage of friends who... (VAS), mean (SD)			
Are sexually active	54 (36)	41 (36)	50 (36)
Engage in risky behavior	25 (27)	20 (27)	23 (27)
Use condoms	67 (36)	46 (35)	60 (37)
Have been tested for HIV	42 (34)	41 (39)	42 (35)

XSL•FO
RenderX

[a]Totals may not sum to N (100%) because of missing data.

[b]LGBTQ: lesbian, gay, bisexual, transgender, or queer.

[c]n=15.

[d]n=45.

[e]n=29.

[f]n=17.

[g]n=13.

[h]n=30.

[i]n=10.

[j]n=3.

[k]n=18.

[l]n=12.

[m]VAS: visual analog scale (0%-100%).

[n]n=14.

[o]n=44.

[p]Reporting on the combined categories of *easy* and *very easy*.

[q]n=19.

HIV Knowledge and Perceptions

Qualitatively, both heterosexual and LGBTQ AYA expressed an advanced understanding of HIV and of the differences between HIV and AIDS (Table 2). An LGBTQ participant explained that HIV "...attacks your immune system" and a heterosexual participant explained that AIDS "...is like the worst...it's like after your T-cells gets to a certain amount then you have AIDS." Both groups of AYA similarly expressed an advanced understanding of HIV prevention strategies and modes of transmission. A heterosexual participant stated of HIV prevention "if you're going to have sex...know your partner...know their status," and an LGBTQ participant stated that HIV could be transmitted through "exchange of fluids." However, both groups also stated myths and misunderstandings about HIV prevention and transmission. Regarding HIV, an LGBTQ participant asked, "...can you get it from urine?" and a heterosexual participant asked, "Is there some kind of vaccine you can take?" In addition, regardless of sexual orientation, both groups held both negative and positive perceptions about HIV. A heterosexual participant shared, "It takes away your sexual life...," whereas an LGBTQ participant felt "you still have a second chance to live...So, if you're still alive I think that you should...say 'I'm still here, I'm still sexy.'" On the basis of the adapted Brief HIV Knowledge Questionnaire, there were only two items for which the average number of participants responding correctly was <50%: an average of 33% (SD 0.474) of the participants knew that HIV infection could not be determined by taking an HIV test 1 week after a potential exposure and an average of 28% (SD 0.455) knew that a vaccine for HIV does not exist (Table 3).

Table 2. Summary of focus group themes related to HIV knowledge.

Themes and subthemes	Representative quotes
HIV knowledge	
General	• "HIV is like you can get treatment, if you get AIDS you could die." [Heterosexual participant]
	• "AIDS is like the worst...it's like after your T-cells gets to a certain amount then you have AIDS." [Heterosexual participant]
	• "...it tears you up, like your immune system and all of that, like it chews you up on the inside." [LGBTQ[a] participant]
Prevention	• "Don't be silly, wrap your willy." [Heterosexual participant]
	• "It's called PrEP[b]." [Heterosexual participant]
	• "Condoms." [LGBTQ participant]
	• "I say don't get drunk and don't get high." [LGBTQ participant]
	• "And you can do PrEP. I think there's this new thing out called PrEP you take it every day I think, it's like a medicine that helps you, yeah." [LGBTQ participant]
Transmission	• "I'm pretty sure it's transmitted sexually, like sexual intercourse." [Heterosexual participant]
	• "Or like sharing needles." [Heterosexual participant]
	• "You can get it from...shooting up [with] dirty needles, getting tattoos with dirty needles...And then you can get it from blood...You can get it from having a lot of unprotected sex...Oh yeah and you can get it from your parents." [LGBTQ participant]
Myths and misunderstandings	• "Mosquitoes, right?" [Heterosexual participant]
	• "AIDS is when you already have an STD[c], and you know next you get infected with HIV like it acts together." [LGBTQ participant]
Perceptions regarding HIV: severity	• "But it's not really a death sentence...some people assume that it's like that. Like when you first hear you have cancer it's like 'oh my gosh I've got cancer I'm about to be dead.' Like people think that it's terminal but it's not like that." [Heterosexual participant]
	• "Okay, so...it's not like a death sentence. People can live with it. But...it's kind of like you feel like your life is ruined afterwards you know you're socially shunned so to speak...You feel like no one's there for you when you really need them..." [LGBTQ participant]

[a]LGBTQ: lesbian, gay, bisexual, transgender, or queer.

[b]PrEP: pre-exposure prophylaxis.

[c]STD: sexually transmitted disease.

Table 3. Individual item means and SDs for the adapted Brief HIV Knowledge Questionnaire among adolescents and young adults.

	Values, mean[a] (SD)
There is a female condom that can help decrease a woman's chance of getting HIV (T[b])	80 (0.401)
People who have been infected with HIV quickly show serious signs of being infected (F[c])	76 (0.431)
Having sex with more than one partner can increase a person's chance of being infected with HIV (T)	69 (0.465)
A person can get HIV from oral sex (T)	63 (0.488)
A person can get HIV by sharing a glass of water with someone who has HIV (F)	61 (0.493)
A person can get HIV by sitting in a hot tub or a swimming pool with a person who has HIV (F)	61 (0.493)
Using Vaseline or baby oil with condoms lowers the chance of getting HIV (F)	61 (0.493)
All pregnant women infected with HIV will have babies born with AIDS (F)	61 (0.493)
Coughing and sneezing DO NOT spread HIV (T)	58 (0.497)
Taking a test for HIV one week after having sex will tell a person if she or he has HIV (F)	33 (0.474)
There is a vaccine that can stop people from getting HIV (F)	28 (0.455)
Average total correct score (of 11 items)	6.52 (2.681)

[a]Mean indicates the average percentage of participants who answered the question correctly.

[b]T: true.

[c]F: false.

XSL•FO

RenderX

Barriers and Facilitators to HIV Testing

During the focus group discussions, the heterosexual and LGBTQ participants described their perceptions of the benefits of HIV testing, prior experiences with HIV testing using various testing modalities, and their perceived barriers and facilitators to HIV testing (Table 4). Regarding HIV testing, a heterosexual participant felt that it should be done even if you are not sexually active: "...I still would say it's best that you do it just so you...get into the habit of getting tested." An LGBTQ participant thought that HIV testing should be offered on college campuses because of the risky behaviors of students: "...'cause I know...campuses...are very risky." When describing their personal experiences with HIV testing, a heterosexual participant explained not knowing that they needed to explicitly opt out: "...I didn't know they tested me. And so, when they took my blood to find out if I had something else...they was like 'everything's negative...HIV negative.'" An LGBTQ participant shared being surprised by the different options for HIV testing: "The first time I did mine it was interesting 'cause I thought you had to get the needle...I got the swab and was kinda surprised."

Regarding perceived barriers to HIV testing, at the individual level, the participants identified fear and a sense of invincibility among AYA. An LGBTQ participant explained: "...with us being so young we believe that we have a whole life ahead of us...some people wouldn't want to get tested because...they want to ignore it." At the interpersonal level, a heterosexual participant cited the lack of parental support as a potential barrier to testing: "...because my mom would be like 'what are you doing that would warrant you to go get tested?'...And so, it's an awkward conversation." At the structural level, an LGBTQ participant explained that if AYA "...don't even know where they can go," they might not get tested for HIV. Regarding perceived facilitators to HIV testing, the participants identified factors at the individual and structural levels. An LGBTQ participant felt that, on an individual level, AYA should be intrinsically motivated to get tested: "It shouldn't have to be a material motive...you should be willing to just go and get tested without something in it for you." On a structural level, a heterosexual participant thought that if commercials and advertisements "...showed where the free clinics are it would help," whereas another LGBTQ participant posited that HIV testing efforts should try "...to appeal to...people using social media...like...Facebook, Twitter, Instagram, and Snapchat...So, like you can definitely submit those sponsored ads to say you know 'hey get tested.'"

Table 4. Summary of focus group themes related to perceived barriers and facilitators to HIV testing.

Themes	Representative quotes
Perceived benefits	• "It allows you to tell if your partner who you're doing things with you can tell them 'Oh, I have HIV' or 'I have this and that' and you can be more conscious about what you're doing." [Heterosexual participant] • "If you get tested right away you won't be able to spread it to other people who are unaware that we had it." [Heterosexual participant] • "...I just prefer like if we talking then I prefer us to go together. Just so it would be like I know what you got, you know what I got and it's no secrets in between." [LGBTQ[a] participant] • "...when I do have sex I think, well I know for sure that I'm always going to get tested because I really don't want no HIV or AIDS." [LGBTQ participant]
Personal experiences	• "They came to us and took me into a private room, and they pricked my finger...And while the resolution was happening, she talked to me about what would happen and everything if it was positive or negative." [Heterosexual participant] • "The first time it was scary because you know it's your first time engaging in sexual intercourse, so you really don't know even though I maintained enough knowledge about it. I still was nervous." [LGBTQ participant]
Perceived barriers to testing	• "A lot of people they just are scared like what if they do have HIV, they probably would rather not know..." [Heterosexual participant] • "There are some places I believe that will charge you. That would make people stay away from it." [Heterosexual participant] • "I don't think there [are] enough commercials and billboards and stuff like that...I always hear about it either from [organization X] or the doctors, the hospital but there's not enough promoting of getting tested." [LGBTQ participant]
Perceived facilitators to testing	• "It's good to have it more accessible like that it shouldn't matter where. Just giving more options." [Heterosexual participant]

[a]LGBTQ: lesbian, gay, bisexual, transgender, or queer.

Acceptability of Game Prototype

In addition to the engaging focus group discussions, brief surveys using Likert scales were used to evaluate acceptability of the game prototype among AYA (Table 5). On a scale of 1 to 5, with 5 indicating strong agreement, most of the participants agreed that the game was interesting (median 5.0, IQR 5.0-5.0), fun (median 5.0, IQR 4.0-5.0), and easy to learn (median 5.0, IQR 5.0-5.0). Most of the participants also liked the game environment (median 5.0, IQR 4.0-5.0), the game interface (median 5.0, IQR 5.0-5.0), the interactions with other characters in the game (median 5.0, IQR 4.0-5.0), the idea of playing games about behaviors that may be associated with HIV infection (median 5.0, IQR 5.0-5.0), and the idea of playing games about different dating scenarios (median 5.0, IQR 5.0-5.0). Participants also agreed that they would share the game with friends to help them get tested for HIV (median 5.0, IQR 4.0-5.0).

Using a separate Likert scale with a range of 0 to 3, with 3 indicating high likelihood, participants reported that they would play the game if it were available to them (median 3.0, IQR 2.0-3.0), they would use this game to help with getting tested for HIV (median 3.0, IQR 2.0-3.0), they would be more likely to get tested for HIV if the game determined their character to be at risk for HIV infection (median 3.0, SD 2.0-3.0), and they would be more likely to get tested if the game connected them with an actual person who could help (median 3.0, IQR 2.0-3.0).

Table 5. Acceptability of game prototype stratified by sexual orientation (N=46).

Item	Heterosexual (n=30), median (IQR)	LGBTQ[a] (n=16), median (IQR)	Total (N=46), median (IQR)
The game playing was very interesting[b]	5.0 (5.0-5.0)	5.0 (5.0-5.0)	5.0 (5.0-5.0)
The game playing was fun[b]	5.0 (5.0-5.0)	5.0 (4.0-5.0)	5.0 (4.0-5.0)
It was easy to learn how to play the game[b]	5.0 (5.0-5.0)	5.0 (5.0-5.0)	5.0 (5.0-5.0)
I liked the art and animation[b]	5.0 (4.0-5.0)	5.0 (4.0-5.0)	5.0 (4.0-5.0)
I liked the game environment[b]	5.0 (4.0-5.0)	4.0 (4.0-5.0)	5.0 (4.0-5.0)
I liked the game interface[b]	5.0 (5.0-5.0)	5.0 (4.0-5.0)	5.0 (5.0-5.0)
I like the interactions with other characters in the game[b]	5.0 (4.0-5.0)	5.0 (4.0-5.0)	5.0 (4.0-5.0)
I like the idea of playing games about behaviors that may be associated with HIV infection[b]	5.0 (5.0-5.0)	5.0 (5.0-5.0)	5.0 (5.0-5.0)
I like the idea of playing games where my character can meet people and make friends[b]	5.0 (5.0-5.0)	5.0 (5.0-5.0)	5.0 (5.0-5.0)
I like the idea of playing games about different dating scenarios[b]	5.0 (5.0-5.0)	5.0 (5.0-5.0)	5.0 (5.0-5.0)
I would share the game with friends or people I know to help them get tested for HIV[b]	5.0 (4.0-5.0)	5.0 (4.0-5.0)	5.0 (4.0-5.0)
I would play these games if I had access to them[c]	3.0 (2.0-3.0)	3.0 (3.0-3.0)	3.0 (2.0-3.0)
I would recommend the games to a friend[c]	3.0 (2.0-3.0)	3.0 (2.0-3.0)	3.0 (2.0-3.0)
I would be interested in playing these games in multiplayer mode with my friends[c]	3.0 (2.0-3.0)	3.0 (2.5-3.0)	3.0 (2.0-3.0)
I would use this game to help with getting tested for HIV[c]	2.0 (2.0-3.0)	3.0 (2.0-3.0)	3.0 (2.0-3.0)
I would be more likely to get tested for HIV if the game determined my character to be at risk for HIV infection[c]	3.0 (2.0-3.0)	3.0 (3.0-3.0)	3.0 (2.0-3.0)
I would be more likely to get tested if the game connected me with an actual person who could help me get tested[c]	3.0 (1.0-3.0)	3.0 (2.0-3.0)	3.0 (2.0-3.0)

[a]LGBTQ: lesbian, gay, bisexual, transgender, or queer.

[b]Responses provided on a scale of 1 to 5: strongly agree (5), somewhat agree (4), neither agree nor disagree (3), disagree (2), strongly disagree (1).

[c]Responses provided on a scale of 0 to 3: very likely (3), somewhat likely (2), not at all likely (1), don't know (0).

Discussion

Principal Findings

The key findings of this study conducted among AYA in DC include a high level of engagement in sexual behaviors that increase the risk of HIV acquisition, a high level of HIV knowledge, and the identification of perceived barriers and facilitators to HIV testing at multiple levels. The findings from the study informed the development of an intervention to increase HIV testing among AYA. As in other studies, AYA in our study engaged in behaviors that might put them at risk for HIV, such as sexual activity and condomless sex [16,17,54,55], with a higher proportion of LGBTQ AYA engaging in these behaviors than heterosexual AYA [3,56]. However, the proportion of AYA in our study having ever been tested for HIV was much higher than that reported in other studies (67% vs 22%-34% in other studies) [14,16,17,54]. This may be the result of DC-wide efforts to increase HIV testing and education among AYA to curb new infections [11,57]. Furthermore, a higher proportion of LGBTQ AYA in our study had been tested for HIV than heterosexual AYA, consistent with other research [55,58].

Compared with AYA in other studies [27-29], the AYA in our study displayed relatively high HIV knowledge, and there were no differences in knowledge between LGBTQ and heterosexual AYA. Again, this may be the result of jurisdiction-wide efforts to increase HIV education among AYA. However, the participants expressed some persistent misconceptions around modes of transmission, the existence of a vaccine, and the appropriate window period for testing after a potential exposure, implying a need for continued education of AYA. The participants also perceived barriers to testing at the individual, interpersonal, and structural level, suggesting a need for

multilevel interventions. Get Connected, one such intervention, is a web-based intervention that aims to reduce individual and structural barriers to HIV and sexually transmitted infection testing and pre-exposure prophylaxis access among YMSM in Philadelphia, Atlanta, and Houston by tailoring content based on individual sociodemographic characteristics, testing history, and sexual behavior and only referring participants to culturally competent, high-quality testing and pre-exposure prophylaxis sites [35].

Limitations

Our study includes several limitations that warrant discussion. First, our study used convenience sampling by recruiting from specific organizations. Thus, our findings may not be generalizable to AYA who do not routinely seek medical care, do not frequent the participating community-based organizations, and are not part of the youth advisory council. The study included a small number of focus group participants, although this is common in qualitative research. Furthermore, most of the participants identified as Black or African American. Although the racial distribution of the study could have been broader to increase the generalizability of the findings, the racial distribution of our study population reflects the high burden of, and risk for, infection among Black AYA in the United States and in DC specifically. Similarly, the generalizability of the study's findings to AYA of other sexual orientations may be limited because most of the AYA in our study identified as heterosexual.

Implications for Research

Achieving the first strategy of the EHE initiative will depend in part on maximizing the perceived benefits of, and minimizing the perceived barriers to, HIV testing among key populations, including AYA. Toward that goal, this study provides useful information because we were able to determine the behaviors potentially placing AYA at risk for HIV, measure baseline HIV knowledge, assess perceived barriers and facilitators to HIV testing, and assess acceptability of the game prototype. These data informed the development of an mHealth intervention, a life-and-dating simulation game, tailored for, and guided by, AYA in an urban area that aims to increase the perceived susceptibility to, and knowledge of, HIV infection of AYA; increase the perceived benefits of HIV testing, motivation to test, and self-efficacy to test; and decrease the perceived barriers to testing among AYA. These aims will be achieved by displaying the consequences of unsafe sexual behavior in real time using the CDC Risk Estimator Tool while also providing a zip code testing locator and empowering messaging around testing to further facilitate access to convenient testing sites [43]. The efficacy of this intervention will be tested in a randomized controlled trial, and the results will, we hope, reduce barriers to HIV testing faced by AYA and further reduce new infections among this key population.

Conflicts of Interest

DG, a principal investigator on the grant, has a stake in Media Rez consistent with the policy on small business innovation research.

Multimedia Appendix 1
Focus group guide.
[DOCX File , 18 KB - formative_v6i3e29196_app1.docx]

References

1. Li Z, Purcell DW, Sansom SL, Hayes D, Hall HI. Vital signs: HIV transmission along the continuum of care - United States, 2016. MMWR Morb Mortal Wkly Rep 2019;68(11):267-272 [FREE Full text] [doi: 10.15585/mmwr.mm6811e1] [Medline: 30897075]

2. Estimated HIV incidence and prevalence in the United States, 2015-2019. Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-supplemental-report-vol-26-1.pdf [accessed 2021-03-29]

3. Kann L, McManus T, Harris WA, Shanklin SL, Flint KH, Queen B, et al. Youth risk behavior surveillance - United States, 2017. MMWR Surveill Summ 2018;67(8):1-114 [FREE Full text] [doi: 10.15585/mmwr.ss6708a1] [Medline: 29902162]

4. Marks G, Crepaz N, Senterfitt JW, Janssen RS. Meta-analysis of high-risk sexual behavior in persons aware and unaware they are infected with HIV in the United States: implications for HIV prevention programs. J Acquir Immune Defic Syndr 2005;39(4):446-453. [doi: 10.1097/01.qai.0000151079.33935.79] [Medline: 16010168]

XSL•FO
RenderX

5. Rotheram-Borus MJ, Futterman D. Promoting early detection of human immunodeficiency virus infection among adolescents. Arch Pediatr Adolesc Med 2000;154(5):435-439. [doi: 10.1001/archpedi.154.5.435] [Medline: 10807291]

6. Weinhardt LS, Carey MP, Johnson BT, Bickham NL. Effects of HIV counseling and testing on sexual risk behavior: a meta-analytic review of published research, 1985-1997. Am J Public Health 1999;89(9):1397-1405. [doi: 10.2105/ajph.89.9.1397] [Medline: 10474559]

7. Diagnoses of HIV infection in the United States and dependent areas, 2019. Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-2018-updated-vol-32.pdf [accessed 2021-03-29]

8. District of Columbia Department of Health. 2020. URL: https://dchealth.dc.gov/sites/default/files/dc/sites/doh/publication/attachments/2020-HAHSTA-Annual-Surveillance-Report-Appendix.pdf [accessed 2021-03-29]

9. District of Columbia Department of Health. 2020. URL: https://dchealth.dc.gov/sites/default/files/dc/sites/doh/publication/attachments/2020-HAHSTA-Annual-Surveillance-Report.pdf [accessed 2021-03-29]

10. What is ending the HIV epidemic in the U.S.? United States Department of Health and Human Services. 2019. URL: https://www.hiv.gov/federal-response/ending-the-hiv-epidemic/overview [accessed 2021-03-29]

11. 90/90/90/50 plan: ending the epidemic in the District of Columbia by 2020. District of Columbia Department of Health. 2016. URL: https://dchealth.dc.gov/page/90909050-plan-ending-epidemic-district-columbia-2020 [accessed 2021-03-29]

12. Youth sexual health plan: District of Columbia 2016-2020. District of Columbia Department of Health. 2016. URL: https://tinyurl.com/3dhkayf3 [accessed 2021-03-29]

13. Branson BM, Viall A, Marum E. Expanding HIV testing: back to the future. J Acquir Immune Defic Syndr 2013;63 Suppl 2:S117-S121. [doi: 10.1097/QAI.0b013e3182986f25] [Medline: 23764622]

14. Inungu J, Lewis A, Mustafa Y, Wood J, O'Brien S, Verdun D. HIV testing among adolescents and youth in the United States: update from the 2009 behavioral risk factor surveillance system. Open AIDS J 2011;5:80-85 [FREE Full text] [doi: 10.2174/1874613601105010080] [Medline: 21915233]

15. Ma M, Malcolm L, Diaz-Albertini K, Klinoff VA. HIV testing characteristics among Hispanic adolescents. J Community Health 2016;41(1):11-14. [doi: 10.1007/s10900-015-0056-7] [Medline: 26093652]

16. Schnall R, Rojas M, Travers J. Understanding HIV testing behaviors of minority adolescents: a health behavior model analysis. J Assoc Nurses AIDS Care 2015;26(3):246-258 [FREE Full text] [doi: 10.1016/j.jana.2014.08.005] [Medline: 25283353]

17. Swenson RR, Rizzo CJ, Brown LK, Payne N, DiClemente RJ, Salazar LF, et al. Prevalence and correlates of HIV testing among sexually active African American adolescents in 4 US cities. Sex Transm Dis 2009;36(9):584-591 [FREE Full text] [doi: 10.1097/OLQ.0b013e3181b4704c] [Medline: 19661840]

18. Van Handel M, Kann L, Olsen EO, Dietz P. HIV testing among US high school students and young adults. Pediatrics 2016;137(2):e20152700. [doi: 10.1542/peds.2015-2700] [Medline: 26787047]

19. Branson BM, Handsfield HH, Lampe MA, Janssen RS, Taylor AW, Lyss SB, Centers for Disease Control and Prevention (CDC). Revised recommendations for HIV testing of adults, adolescents, and pregnant women in health-care settings. MMWR Recomm Rep 2006;55(RR-14):1-17 [FREE Full text] [Medline: 16988643]

20. US Preventive Services Task Force, Owens DK, Davidson KW, Krist AH, Barry MJ, Cabana M, et al. Screening for HIV infection: US preventive services task force recommendation statement. JAMA 2019;321(23):2326-2336. [doi: 10.1001/jama.2019.6587] [Medline: 31184701]

21. Committee on Pediatric AIDS, Emmanuel PJ, Martinez J. Adolescents and HIV infection: the pediatrician's role in promoting routine testing. Pediatrics 2011;128(5):1023-1029. [doi: 10.1542/peds.2011-1761] [Medline: 22042816]

22. Dimmitt Champion J, Harlin B, Collins JL. Sexual risk behavior and STI health literacy among ethnic minority adolescent women. Appl Nurs Res 2013;26(4):204-209 [FREE Full text] [doi: 10.1016/j.apnr.2013.06.003] [Medline: 23867137]

23. Goyal MK, Witt R, Hayes KL, Zaoutis TE, Gerber JS. Clinician adherence to recommendations for screening of adolescents for sexual activity and sexually transmitted infection/human immunodeficiency virus. J Pediatr 2014;165(2):343-347 [FREE Full text] [doi: 10.1016/j.jpeds.2014.04.009] [Medline: 24840761]

24. Hyden C, Allegrante JP, Cohall AT. HIV testing sites' communication about adolescent confidentiality: potential barriers and facilitators to testing. Health Promot Pract 2014;15(2):173-180. [doi: 10.1177/1524839913499347] [Medline: 23966274]

25. Peralta L, Deeds BG, Hipszer S, Ghalib K. Barriers and facilitators to adolescent HIV testing. AIDS Patient Care STDS 2007;21(6):400-408. [doi: 10.1089/apc.2006.0112] [Medline: 17594249]

26. Gurvey J, Adler N, Ellen JM. Factors associated with self-risk perception for sexually transmitted diseases among adolescents. Sex Transm Dis 2005;32(12):742-744. [doi: 10.1097/01.olq.0000175385.23447.27] [Medline: 16314770]

27. Míguez MJ, Espinoza LA, Vargas ME, Perez C, Ergon E, Tarter R. Low HIV/AIDS knowledge among Hispanic adolescents. J AIDS Clin Res 2015;6(7):483 [FREE Full text] [doi: 10.4172/2155-6113.1000483] [Medline: 30627475]

28. Swenson RR, Rizzo CJ, Brown LK, Vanable PA, Carey MP, Valois RF, et al. HIV knowledge and its contribution to sexual health behaviors of low-income African American adolescents. J Natl Med Assoc 2010;102(12):1173-1182 [FREE Full text] [doi: 10.1016/s0027-9684(15)30772-0] [Medline: 21287898]

29. Udell W, Donenberg G. Remembering the basics: African American youth and HIV knowledge - brief report. J Natl Med Assoc 2015;107(3):20-24. [doi: 10.1016/S0027-9684(15)30047-X] [Medline: 27282719]

30. Murphy DA, Mitchell R, Vermund SH, Futterman D, Adolescent Medicine HIV/AIDS Research Network. Factors associated with HIV testing among HIV-positive and HIV-negative high-risk adolescents: the REACH study. Reaching for Excellence in Adolescent Care and Health. Pediatrics 2002;110(3):e36. [doi: 10.1542/peds.110.3.e36] [Medline: 12205286]

31. 2019 youth risk behavior survey results: District of Columbia middle school survey: risk behaviors and sexual identity report. District of Columbia Office of the State Superintendent of Education. 2019. URL: https://osse.dc.gov/sites/default/files/dc/sites/osse/page_content/attachments/2019DCBM%20Sexual%20Identity.pdf [accessed 2021-03-29]

32. 2019 youth risk behavior survey results: District of Columbia high school survey: risk behaviors and sexual identity report. District of Columbia Office of the State Superintendent of Education. 2019. URL: https://osse.dc.gov/sites/default/files/dc/sites/osse/page_content/attachments/2019DCBH%20Sexual%20Identity.pdf [accessed 2021-03-29]

33. Dowshen N, Lee S, Matty Lehman BM, Castillo M, Mollen C. IknowUshould2: feasibility of a youth-driven social media campaign to promote STI and HIV testing among adolescents in Philadelphia. AIDS Behav 2015;19 Suppl 2(0 2):106-111 [FREE Full text] [doi: 10.1007/s10461-014-0991-9] [Medline: 25563502]

34. McFarlane M, Brookmeyer K, Friedman A, Habel M, Kachur R, Hogben M. GYT: get yourself tested campaign awareness: associations with sexually transmitted disease/HIV testing and communication behaviors among youth. Sex Transm Dis 2015;42(11):619-624 [FREE Full text] [doi: 10.1097/OLQ.0000000000000361] [Medline: 26457487]

35. Bauermeister JA, Golinkoff JM, Horvath KJ, Hightow-Weidman LB, Sullivan PS, Stephenson R. A multilevel tailored web app-based intervention for linking young men who have sex with men to quality care (get connected): protocol for a randomized controlled trial. JMIR Res Protoc 2018;7(8):e10444 [FREE Full text] [doi: 10.2196/10444] [Medline: 30072358]

36. Mustanski B, Parsons JT, Sullivan PS, Madkins K, Rosenberg E, Swann G. Biomedical and behavioral outcomes of keep it up!: an eHealth HIV prevention program RCT. Am J Prev Med 2018;55(2):151-158 [FREE Full text] [doi: 10.1016/j.amepre.2018.04.026] [Medline: 29937115]

37. Kachur R, Mesnick J, Liddon N, Kapsimalis C, Habel M, David-Ferdon C, et al. Adolescents, technology and reducing risk for HIV, STDs, and pregnancy. Centers for Disease Control and Prevention. 2013. URL: https://www.cdc.gov/std/life-stages-populations/adolescents-white-paper.pdf [accessed 2021-03-29]

38. Castel AD, Qasmieh S, Greenberg D, Ellenberger N, Howell TH, Griffith C, et al. Digital gaming to improve adherence among adolescents and young adults living with HIV: mixed-methods study to test feasibility and acceptability. JMIR Serious Games 2018;6(4):e10213 [FREE Full text] [doi: 10.2196/10213] [Medline: 30322838]

39. Cordova D, Alers-Rojas F, Lua FM, Bauermeister J, Nurenberg R, Ovadje L, Youth Leadership Council. The usability and acceptability of an adolescent mHealth HIV/STI and drug abuse preventive intervention in primary care. Behav Med 2018;44(1):36-47 [FREE Full text] [doi: 10.1080/08964289.2016.1189396] [Medline: 27223646]

40. Jones J, Dominguez K, Stephenson R, Stekler JD, Castel AD, Mena LA, et al. A theoretically based mobile app to increase pre-exposure prophylaxis uptake among men who have sex with men: protocol for a randomized controlled trial. JMIR Res Protoc 2020;9(2):e16231 [FREE Full text] [doi: 10.2196/16231] [Medline: 32130178]

41. Kuhns LM, Garofalo R, Hidalgo M, Hirshfield S, Pearson C, Bruce J, et al. A randomized controlled efficacy trial of an mHealth HIV prevention intervention for sexual minority young men: MyPEEPS mobile study protocol. BMC Public Health 2020;20(1):65 [FREE Full text] [doi: 10.1186/s12889-020-8180-4] [Medline: 31941475]

42. Pendergrass T, Hieftje K, Duncan L, Fiellin L. Videogame intervention to encourage HIV testing and counseling among adolescents. Mhealth 2020;6:26 [FREE Full text] [doi: 10.21037/mhealth.2020.01.05] [Medline: 32632364]

43. Wilbourn B, Howell TH, Castel AD, D'Angelo L, Trexler C, Carr R, et al. Development, refinement, and acceptability of digital gaming to improve HIV testing among adolescents and young adults at risk for HIV. Games Health J 2020;9(1):53-63 [FREE Full text] [doi: 10.1089/g4h.2018.0162] [Medline: 31560218]

44. Janz NK, Becker MH. The health belief model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]

45. Dorr N, Krueckeberg S, Strathman A, Wood MD. Psychosocial correlates of voluntary HIV antibody testing in college students. AIDS Educ Prev 1999;11(1):14-27. [Medline: 10070586]

46. Povinelli M, Remafedi G, Tao G. Trends and predictors of human immunodeficiency virus antibody testing by homosexual and bisexual adolescent males, 1989-1994. Arch Pediatr Adolesc Med 1996;150(1):33-38. [doi: 10.1001/archpedi.1996.02170260037005] [Medline: 8542003]

47. Rosenstock IM, Strecher VJ, Becker MH. The health belief model and HIV risk behavior change. In: DiClemente RJ, Peterson JL, editors. Preventing AIDS: theories and methods of behavioral interventions. New York, NY: Plenum Press; 1994:6-24.

48. Thomas R, Cahill J, Santilli L. Using an interactive computer game to increase skill and self-efficacy regarding safer sex negotiation: field test results. Health Educ Behav 1997;24(1):71-86. [doi: 10.1177/109019819702400108] [Medline: 9112099]

49. Guest G, Bunce A, Johnson L. How many interviews are enough?: an experiment with data saturation and variability. Field Methods 2006;18(1):59-82. [doi: 10.1177/1525822X05279903]

50. Heckman TG, Kelly JA, Sikkema K, Cargill V, Norman A, Fuqua W, et al. HIV risk characteristics of young adult, adult, and older adult women who live in inner-city housing developments: implications for prevention. J Womens Health (Larchmt) 1995;4(4):397-406. [doi: 10.1089/jwh.1995.4.397]

51. Healthy Oakland teens: healthy Oakland teens description and explanation of study instrument. University of California San Francisco Department of Medicine. URL: https://prevention.ucsf.edu/research-project/healthy-oakland-teens-0 [accessed 2021-03-29]

52. Carey MP, Schroder KE. Development and psychometric evaluation of the brief HIV knowledge questionnaire. AIDS Educ Prev 2002;14(2):172-182 [FREE Full text] [doi: 10.1521/aeap.14.2.172.23902] [Medline: 12000234]

53. Charmaz K. Grounded theory. In: Smith JA, editor. Qualitative psychology: a practical guide to research methods. Thousand Oaks, CA: Sage Publications; 2007.

54. Coeytaux K, Kramer MR, Sullivan PS. HIV testing among United States high school students at the state and national level, Youth Risk Behavior Survey 2005-2011. Springerplus 2014;3:202 [FREE Full text] [doi: 10.1186/2193-1801-3-202] [Medline: 24855587]

55. Straub DM, Arrington-Sanders R, Harris DR, Willard N, Kapogiannis B, Emmanuel P, Adolescent Trials Network for HIV/AIDS Interventions. Correlates of HIV testing history among urban youth recruited through venue-based testing in 15 US cities. Sex Transm Dis 2011;38(8):691-696 [FREE Full text] [doi: 10.1097/OLQ.0b013e318214bb70] [Medline: 21758020]

56. Mojola SA, Everett B. STD and HIV risk factors among U.S. young adults: variations by gender, race, ethnicity and sexual orientation. Perspect Sex Reprod Health 2012;44(2):125-133 [FREE Full text] [doi: 10.1363/4412512] [Medline: 22681428]

57. Rakhmanina N, Messenger N, Phillips 2nd G, Teach S, Morrison S, Hern J, et al. Factors affecting acceptance of routine human immunodeficiency virus screening by adolescents in pediatric emergency departments. J Adolesc Health 2014;54(2):176-182. [doi: 10.1016/j.jadohealth.2013.07.027] [Medline: 24060572]

58. Adebayo OW, Gonzalez-Guarda RM. Factors associated with HIV testing in youth in the United States: an integrative review. J Assoc Nurses AIDS Care 2017;28(3):342-362. [doi: 10.1016/j.jana.2016.11.006] [Medline: 27993497]

Abbreviations

AYA: adolescents and young adults
CDC: Centers for Disease Control and Prevention
DC: Washington, District of Columbia
EHE: Ending the HIV Epidemic in the U.S.
LGBTQ: lesbian, gay, bisexual, transgender, or queer
mHealth: mobile health
YMSM: young men who have sex with men
YRBS: Youth Risk Behavior Survey

XSL•FO
RenderX

Original Paper

Searching for HIV and AIDS Health Information in South Africa, 2004-2019: Analysis of Google and Wikipedia Search Trends

Babatunde Okunoye[1,2], BA; Shaoyang Ning[3], PhD; Dariusz Jemielniak[4], PhD

[1]Berkman Klein Centre for Internet and Society, Harvard University, Cambridge, MA, United States

[2]Department of Journalism, Film and Television, University of Johannesburg, Johannesburg, South Africa

[3]Department of Mathematics and Statistics, Williams College, Massachusetts, MA, United States

[4]Management in Networked and Digital Societies Department, Kozminski University, Warsaw, Poland

Corresponding Author:
Babatunde Okunoye, BA
Berkman Klein Centre for Internet and Society
Harvard University
2nd Floor
23 Everett Street
Cambridge, MA, 02138
United States
Phone: 1 (617) 495 7547
Email: bokunoye@cyber.harvard.edu

Abstract

Background: AIDS, caused by HIV, is a leading cause of mortality in Africa. HIV/AIDS is among the greatest public health challenges confronting health authorities, with South Africa having the greatest prevalence of the disease in the world. There is little research into how Africans meet their health information needs on HIV/AIDS online, and this research gap impacts programming and educational responses to the HIV/AIDS pandemic.

Objective: This paper reports on how, in general, interest in the search terms "HIV" and "AIDS" mirrors the increase in people living with HIV and the decline in AIDS cases in South Africa.

Methods: Data on search trends for HIV and AIDS for South Africa were found using the search terms "HIV" and "AIDS" (categories: health, web search) on Google Trends. This was compared with data on estimated adults and children living with HIV, and AIDS-related deaths in South Africa, from the Joint United Nations Programme on HIV/AIDS, and also with search interest in the topics "HIV" and "AIDS" on Wikipedia Afrikaans, the most developed local language Wikipedia service in South Africa. Nonparametric statistical tests were conducted to support the trends and associations identified in the data.

Results: Google Trends shows a statistically significant decline ($P<.001$) in search interest for AIDS relative to HIV in South Africa. This trend mirrors progress on the ground in South Africa and is significantly associated ($P<.001$) with a decline in AIDS-related deaths and people living longer with HIV. This trend was also replicated on Wikipedia Afrikaans, where there was a greater interest in HIV than AIDS.

Conclusions: This statistically significant ($P<.001$) association between interest in the search terms "HIV" and "AIDS" in South Africa (2004-2019) and the number of people living with HIV and AIDS in the country (2004-2019) might be an indicator that multilateral efforts at combating HIV/AIDS—particularly through awareness raising and behavioral interventions in South Africa—are bearing fruit, and this is not only evident on the ground, but is also reflected in the online information seeking on the HIV/AIDS pandemic. We acknowledge the limitation that in studying the association between Google search interests on HIV/AIDS and cases/deaths, causal relationships should not be drawn due to the limitations of the data.

(JMIR Form Res 2022;6(3):e29819)   doi:10.2196/29819

XSL•FO
RenderX

Introduction

A major obstacle to combating the impacts of disease in developing countries is the paucity of high-quality health data, including data regarding people's health information needs [1]. If they do not understand people's everyday concerns, health organizations and policy makers are less able to effectively target education and programming efforts for all genders and age groups [1]. People's information needs and their everyday concerns are often expressed via search engine queries as millions go online to meet their health information needs.

HIV/AIDS remains a major global public health concern, despite concerted efforts aimed at combating the disease. There are an estimated 37.9 (32.7-44.0) million people worldwide living with HIV, the virus that causes AIDS [2]. Sub-Saharan Africa has the greatest burden of HIV/AIDS, with South Africa bearing the greatest burden of the disease in the world. An estimated 7.7 million people were living with HIV and AIDS in South Africa in 2018, with an adult prevalence rate of 20.4% [2]. In 2018, approximately 71,000 South Africans died of AIDS-related causes.

Considerable progress has been made in the fight against HIV/AIDS globally. Globally, an estimated 37.9 (32.7-44.0) million people were living with HIV in 2018 [2], an increase from previous years (eg, 35.3 million in 2012), as more people receive life-saving antiretroviral therapy. At the same time, the number of AIDS-related deaths is also declining, with 770,000 deaths in 2018 [2], down from 1.6 (1.4-1.9) million deaths in 2012 and 2.3 (2.1-2.6) million in 2005 [3]. Gains have also been made toward many of the 2020 Sustainable Development Goals targets and elimination commitments, although significant challenges remain [4]. This is within the context of high-risk behavior among people living with HIV in low- and middle-income countries [5] and insufficient human resources in HIV care in low- and middle-income countries [6].

A key component of the fight against HIV/AIDS has been behavioral interventions [3]. A global meta-analysis of studies determined that behavioral interventions reduce sexual risk behavior and reduce sexually transmitted infections and HIV [3]. Behavioral interventions typically involve harmonizing messages and the dissemination of information about HIV transmission and various prevention approaches using television, radio, outdoor advertising, and information and communications technology. The outcome of all these behavioral interventions is that in sub-Saharan Africa, the percentage of young people (15-24 years) demonstrating a comprehensive and accurate understanding of HIV has generally increased over the years. For example, this understanding rose by 5 percentage points for men and by 3 for women from 2002 to 2011, although knowledge levels remain low (36% for young men and 28% for young women) [3].

In South Africa, there have been a number of nationwide communication campaigns related to raising awareness of HIV and AIDS. Soul City and MTV Shuga are two examples of such multimedia campaigns that promote good sexual health and well-being and have effective outcomes [7]. Research into the impact of the Soul City and other communication campaigns found that they had a positive effect on the sexual behavior of adults that had been exposed to the campaign message, for instance by bringing about positive changes in condom use and HIV testing [8].

The combined effects of these awareness campaigns have been an increase in knowledge and understanding of HIV/AIDS. One likely outcome of an increase in awareness of HIV/AIDS in South Africa is that a percentage of the population will turn to the internet via search engines for information on HIV/AIDS. Internet penetration in South Africa is 56% (32.9 million) [9].

An increasing number of people are using the internet to support their health care needs [10,11] and search engines such as Google, Yahoo! and Bing are often their first port of call. One of the early studies on search engine use found that 1 in every 28 (3.5%) pages viewed on the web is a search results page, making the use of a search engine the second most popular internet task next to email [12]. Another early study found that there are over 6.75 million health-related internet searches carried out every day worldwide, representing approximately 5% of all internet searches [13]. Search engine use is the most common approach to online information seeking [14], and a study by the Pew Research Center found that half of all internet users now use search engines on a typical day [15]. It has also been estimated that about 1 billion Google searches daily are health-related, which is about 7% of all daily searches [16].

This paper examines trends in online search on HIV and AIDS on Google, which is the dominant search engine worldwide [17] and is the top search engine in most countries with a few exceptions, such as Russia (Yandex) and China (Baidu). In South Africa, Google is also the dominant search engine [17]. Google search data have been used for understanding health information needs, health surveillance, and health forecasting [10,18-23]. Although the use of Google search engine data for health and development in general has been criticized [24,25] in recent years due to an unbalanced and opaque approach that did not satisfy scientific transparency, in recent years, organizations such as the World Bank [26,27] and several government departments around the world have begun revisiting the behavioral big data methods of search engine data in aid of international development. Drawing on feedback from the scientific community on the best practices of big data research [24,25,28,29], particularly paying attention to integrating both big data and "small data" (eg, survey data and other traditional data collection mechanisms) to create more complete and accurate data sets, they have made significant advances in this field. This paper advances the hypothesis that the online search behavior of people looking for health information can be used to understand their health information needs, and also for health surveillance and monitoring [10,18-23].

Methods

The methods employed in this study were derived from the considerable literature regarding using search data to understand the health information needs of the public and for health surveillance [10,18-23]. The first step involved search term selection. The search terms chosen as proxies to gauge public interest in HIV and AIDS in South Africa were "HIV" and

"AIDS." These were the only search terms selected given the specificity of the topics concerned. Data [30] on online searches for the terms "HIV" and "AIDS" were thus obtained on Google Trends (categories: health, web search) to visualize how searches were conducted on both search terms. The data were obtained for the years 2004-2019. Some of the most important search queries (top and rising) searched for by South Africans on "HIV" and "AIDS" were downloaded from the Google Trends data set [30] and recorded. Data on regional interest in both search terms for South Africa were downloaded. In addition, the top and rising search queries related to "HIV" and "AIDS" were also accessed via Google Trends.

The nonparametric Mann-Kendall test [31] was conducted to evaluate the significance of trends. The Mann-Kendall test accounts for the autocorrelation of time series (for the alternative hypothesis of monotonic trends) by the design of its test statistic on the time series increments. The Mann-Kendall test was adapted based on the log ratio of search volumes between search terms to evaluate their associations (between "HIV" and "AIDS") and to investigate the search trend of the term "AIDS" with adjustment (log ratio between "flu" and "AIDS").

Nonparametric Mann-Kendall tests were conducted to indicate the significant decline in AIDS-related deaths in South Africa and the increasing trend for the number of estimated adults and children living with HIV. Permutation test based on nonparametric Spearman rho [32] was conducted to verify the association between Google search interest for the term "AIDS" and AIDS-related deaths in Africa. Due to the difference in time frequency, monthly Google search data for the term "AIDS" were aggregated through moving average to estimate yearly search interest. All statistical analysis was conducted with R (version 4.1.2; R Foundation for Statistical Computing); Mann-Kendall tests were done using the *trend* package and the Spearman test was performed using the *stats* package.

Pageviews for the topics "HIV" and "AIDS" from 2015-2020 in Wikipedia Afrikaans, the most developed indigenous language Wikipedia service in South Africa, were obtained from Wikipedia [33]. The year with the earliest data on HIV and AIDS on Wikipedia Afrikaans is 2015.

Review studies on the use of search engine data to understand public interests and trends have noted the "traps in big data" [24,25,28,29], emphasizing that data sources such as search engines are best used to supplement traditional data sources like surveys and that, in general, confidence in such data is strengthened when their findings can be replicated using other data sources and platforms. Thus, this paper also obtained and plotted data [34] on the estimated number of adults and children living with HIV, as well as AIDS-related deaths in South Africa, from the United Nations Joint Programme on HIV/AIDS for the years 2004-2019.

Results

Search term trends for "HIV" and "AIDS" from 2004-2019 in South Africa showed a decline in "AIDS" searches relative to "HIV" ($P<.001$; Figure 1). This observed online behavior mirrors progress on the ground for both HIV and AIDS ($P<.001$, association between "AIDS" searches and deaths), as there was a decrease in AIDS-related deaths ($P<.001$) and an increase in the number of people living with HIV ($P<.001$) in the country from 2004-2019 (Figure 2).

Some of the most important search queries (top and rising) searched for by South Africans on "HIV" and "AIDS" can be seen in Multimedia Appendix 1 and Multimedia Appendix 2, respectively. A complete visualization of the search data, regional interest in South Africa for the search terms "HIV" and "AIDS," and rising and top search queries on "HIV" and "AIDS" can be accessed via Google Trends [30].

Similarly, on Wikipedia Afrikaans, for the years 2015-2020, there was greater interest in "HIV" than "AIDS" (Multimedia Appendix 3).

Figure 1. Search interest in the search terms "HIV" and "AIDS" (category: health, web search) in South Africa, 2004-2019 (accessed December 24, 2020).

Figure 2. UNAIDS data on estimated adults and children living with HIV and AIDS-related deaths in South Africa [34]. UNAIDS: Joint United Nations Programme on HIV/AIDS.
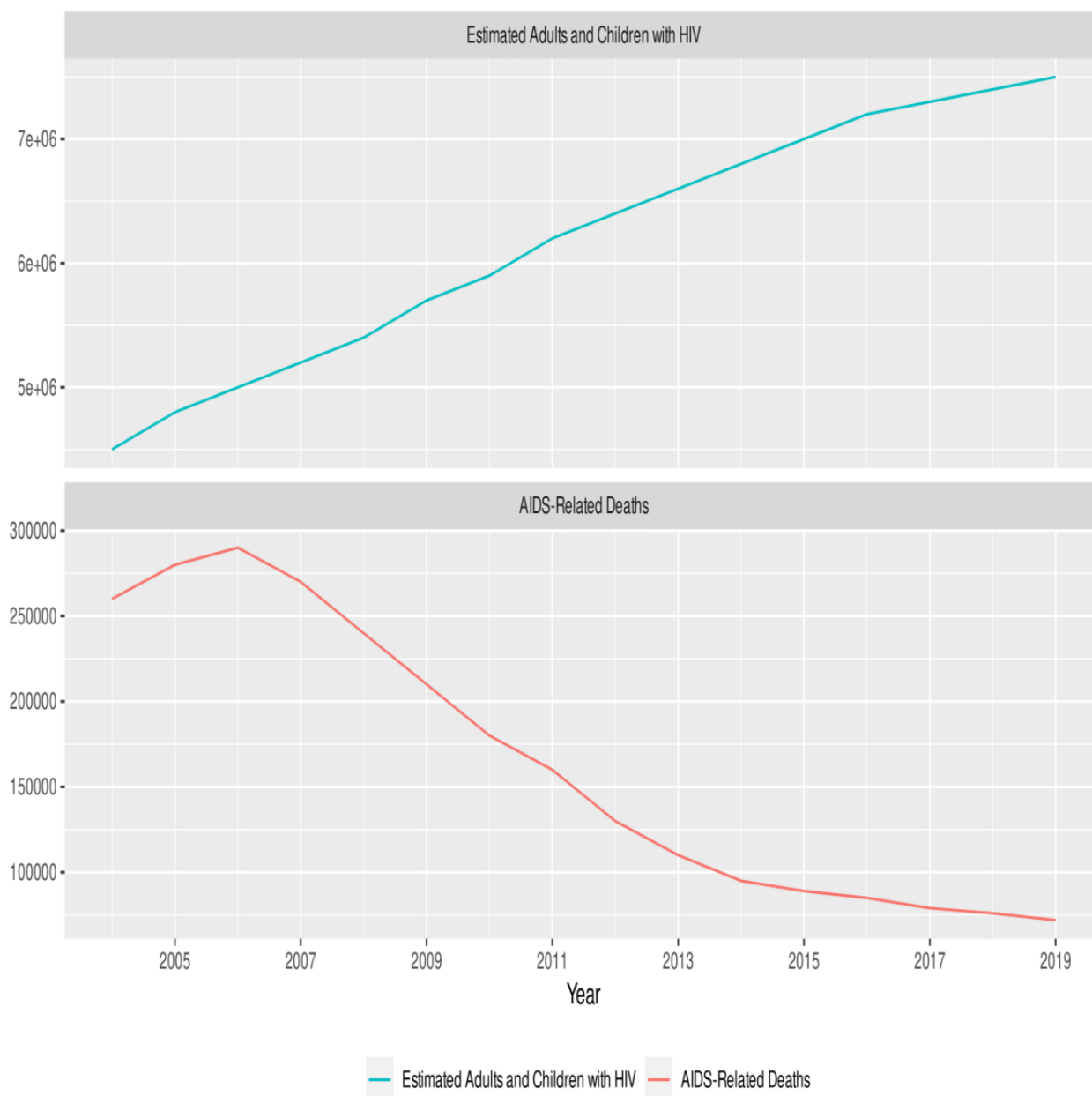
Discussion

Principal Findings

This paper shows a statistically significant decline ($P<.001$) in search interest on Google for AIDS relative to HIV in South Africa. This trend mirrors progress on the ground in South Africa and is significantly associated ($P<.001$) with a decline in AIDS-related deaths and people living longer with HIV. This trend was also replicated on Wikipedia Afrikaans, where there was a greater interest in HIV than AIDS. The data on the search terms "HIV" and "AIDS" in South Africa (2004-2019) reveal a decrease in interest over time in AIDS-related search queries relative to HIV-related search queries. This mirrors progress on the ground in South Africa, where AIDS-related deaths have reduced over the years (2004-2019) and people are living longer with HIV due to life-changing medications. Similarly, behavioral

interventions through messages delivered through media sources in South Africa, including the internet, have increased the number of people with an accurate understanding of the differences between HIV and AIDS [3]. In response to these behavioral interventions, a proportion of the populations in these countries will use the internet via search engines to seek information on HIV and AIDS. From the earliest data available on Wikipedia Afrikaans (2015-2020), there has been a greater interest in HIV than AIDS in South Africa.

Comparison to Prior Work

This work contributes to the growing literature on infodemiology, which seeks to employ search engine data in the service of health information seeking, forecasting, and surveillance [35]. Previous research has covered a wide variety of infectious diseases such as influenza [36], dengue [37],

norovirus [38], COVID-19 [39], and sexually transmitted infections [40], as well as other topics like cancer [22,23], dementia [18], and mental health [20]. Among the limited literature exploring HIV/AIDS with internet search data, the methodology is often restricted to qualitative [41] or parametric methodology [42], with limited research on Africa. Our work contributes to the limited literature (eg, [1,43]) on HIV/AIDS in the African context. Moreover, the nonparametric methodology we adopted requires minimal assumptions about the data and its distribution, and is thus robust and adaptive. Our work contributes to the niche of infodemiological study on HIV/AIDS in Africa and South Africa in particular while possessing the potential to extend to other countries/regions or geographical resolutions.

Further Directions

An important follow-up to this study might be to understand how South Africans search for health information on other diseases and topics of public interest on search engines. In addition, motivated by our findings in this work, we may further investigate and develop a statistically principled and coherent framework to use Google search data for forecasting HIV/AIDS trends. In the meantime, a quantitative (survey) and qualitative (interviews) study of how South Africans access HIV/AIDS information online is planned, to provide insights for the research question by using traditional research methods. This study aims to provide data across South African provinces, disaggregated by age and gender. Additionally, broadening the scope of future studies could be achieved by studying other search engines, social media, and knowledge repositories, as research shows that, for example, Wikipedia is a good reflection of medical interests of the lay population [44], search engine queries are a good way of estimating outbreaks [45], and social media platforms are an important source of medical information today [46].

Limitations

This paper reports on how, in general, interest in the search terms "HIV" and "AIDS" mirrors the increase in people living with HIV and the decline in AIDS cases in South Africa. Here, we report some limitations of the study.

First, there is an acknowledgement that the population of health information seekers on HIV/AIDS online may be quite different from the offline population. For instance, not everyone searching for information on HIV/AIDS may be connected online or use search engines, and the number of people connected to the internet changed over the years of the study (2004-2019). Hence, this paper is not an exact mapping of the online behavior of all the people searching for HIV/AIDS information in South Africa. This limitation is common in research using online big data methods such as search engine data, where the nature of the specific platform from which the data were obtained shapes the results obtained [25]. To mitigate this, a follow-up quantitative

and qualitative study of the HIV/AIDS health information seeking behavior of South Africans using Google is planned. This would provide data through traditional (non–big data) research methods, following best practices from digital big data research, where the best results are often obtained when traditional data collection methods like surveys are combined with big data methods to inform research inquiry [25].

Second, we also acknowledge that the observed trends in the Google search data could be explained by the general decrease in people's interest in such diseases and public health matters. To adjust for this, we used the search term "flu" as a control and the declining trend of the search term "AIDS" remained significant after adjustment ($P<.001$). In addition, this paper acknowledges the effects of media coverage on interest in specific terms, which is not accounted for in this paper.

Third, this paper focuses only on studying the association between the Google search interest in HIV/AIDS and cases/deaths. We acknowledge that causal relationships should not be drawn due to the limitations of the data. In addition, we acknowledge that although we used the Google Trends category "health, web search" to refine the search results for "AIDS," the returned results may contain other meanings or associations of the word "AIDS," as seen in some entries in Multimedia Appendix 2.

Conclusions

A major obstacle to combating the impacts of disease in developing countries is the paucity of high-quality health data, particularly for understanding people's health information needs [1]. People's information needs and their everyday concerns are often expressed via search engine queries as millions go online to meet their health information needs. This paper shows a statistically significant decline ($P<.001$) in search interest on Google for AIDS relative to HIV in South Africa. This trend mirrors progress on the ground in South Africa and is significantly associated ($P<.001$) with a decline in AIDS-related deaths and people living longer with HIV. This trend was also replicated on Wikipedia Afrikaans, where there was a greater interest in HIV than AIDS.

In developing country contexts where high-quality data on the health information needs of people is often lacking, understanding the population's use of search engines to meet health information needs can provide useful data. Investigating people's use of search engines to meet information needs can also reveal if education and programming efforts have been effective. Consequently, a natural progression of this study might be a quantitative and qualitative study of the use of search engines and other online information sources to meet HIV- and AIDS-related health information needs in South Africa, which might mitigate platform-specific limitations of findings as outlined in the Limitations section.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1
Top and rising related queries for HIV, 2004-2019.
[DOCX File , 14 KB - formative_v6i3e29819_app1.docx]

Multimedia Appendix 2
Top and rising related queries for AIDS, 2004-2019.
[DOCX File , 13 KB - formative_v6i3e29819_app2.docx]

Multimedia Appendix 3
Wikipedia Afrikaans pageview statistics on HIV and AIDS 2015-2020 (accessed November 5, 2020).
[DOCX File , 151 KB - formative_v6i3e29819_app3.docx]

References

1. Abebe R, Hill S, Vaughan J, Small P, Schwartz H. Using Search Queries to Understand Health Information Needs in Africa. In: Proceedings of the Thirteenth International AAAI Conference on Web and Social Media. Burnaby: PKP Publishing Services Network; 2019 Presented at: Thirteenth International AAAI Conference on Web and Social Media. Association for the Advancement of Artificial Intelligence; June 11-14, 2019; Munich, Germany URL: https://ojs.aaai.org/index.php/ICWSM/article/view/3360/3228

2. UNAIDS data 2019. UNAIDS. URL: https://www.unaids.org/en/resources/documents/2019/2019-UNAIDS-data [accessed 2020-10-31]

3. 2012 UNAIDS Report on the Global AIDS Epidemic. UNAIDS. URL: https://www.unaids.org/en/resources/documents/2012/20121120_UNAIDS_Global_Report_2012 [accessed 2020-10-31]

4. The Sustainable Development Goals and the HIV response. UNAIDS. 2017. URL: https://www.unaids.org/sites/default/files/media_asset/SDGsandHIV_en.pdf [accessed 2020-10-31]

5. Nguyen PT, Gilmour S, Le PM, Nguyen TT, Tanuma J, Nguyen HV. Factors associated with high-risk behaviors of people newly diagnosed with HIV/AIDS: results from a cross-sectional study in Vietnam. AIDS Care 2021 May;33(5):607-615. [doi: 10.1080/09540121.2020.1761519] [Medline: 32397740]

6. Nguyen PT, Nguyen HV, Phan HT, Cao CT. Situation of human resources and staff training of four provincial HIV/AIDS control centers in Northern Vietnam and some related factors in 2013. VPJM 2015;7(167):75-82.

7. HIV and AIDS in South Africa. Avert. 2020. URL: https://www.avert.org/professionals/hiv-around-world/sub-saharan-africa/south-africa [accessed 2022-01-20]

8. Peltzer K, Parker W, Mabaso M, Makonko E, Zuma K, Ramlagan S. Impact of national HIV and AIDS communication campaigns in South Africa to reduce HIV risk behaviour. ScientificWorldJournal 2012;2012:384608-384606 [FREE Full text] [doi: 10.1100/2012/384608] [Medline: 23213285]

9. Individuals using the Internet 2000-2018. International Telecommunications Union. 2019. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2019/Individuals_Internet_2000-2018_Dec2019.xls [accessed 2020-10-31]

10. Pehora C, Gajaria N, Stoute M, Fracassa S, Serebale-O'Sullivan R, Matava CT. Are Parents Getting it Right? A Survey of Parents' Internet Use for Children's Health Care Information. Interact J Med Res 2015 Jun 22;4(2):e12 [FREE Full text] [doi: 10.2196/ijmr.3790] [Medline: 26099207]

11. Massey PM. Where Do U.S. Adults Who Do Not Use the Internet Get Health Information? Examining Digital Health Information Disparities From 2008 to 2013. J Health Commun 2016;21(1):118-124. [doi: 10.1080/10810730.2015.1058444] [Medline: 26166484]

12. Jansen BJ, Pooch U. A review of Web searching studies and a framework for future research. Journal of the American Society for Information Science and Technology 2001;52(3):235-246. [doi: 10.1002/1097-4571(2000)9999:9999<::AID-ASI1607>3.0.CO;2-F]

13. Eysenbach G, Kohler C. What is the prevalence of health-related searches on the World Wide Web? Qualitative and quantitative analysis of search engine queries on the internet. AMIA Annu Symp Proc 2003:225-229 [FREE Full text] [Medline: 14728167]

14. Morris M, Teevan J, Panovich K. A comparison of Information seeking on search engines and social networks. In: Proceedings of the 4th International AAAI Conference on Weblogs and Social Media. Menlo Park, California: AAAI; 2010 Presented at: 4th International AAAI Conference on Weblogs and Social Media; May 23-26; Washington DC, USA p. 1-4 URL: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/icwsm10.pdf

15. Fallows D. Search Engine Use. Pew Research Center. 2008. URL: https://www.pewresearch.org/internet/2008/08/06/search-engine-use/ [accessed 2020-10-31]

16. Murphy M. Dr Google will see you now: Search giant wants to cash in on your medical queries. The Telegraph. 2019. URL: https://www.telegraph.co.uk/technology/2019/03/10/google-sifting-one-billion-health-questions-day/ [accessed 2021-05-27]

17. Search Engine Market Share. StatCounter Global Stats. URL: https://gs.statcounter.com/search-engine-market-share [accessed 2020-10-30]

18. Wang H, Chen D, Yu H, Chen Y. Forecasting the Incidence of Dementia and Dementia-Related Outpatient Visits With Google Trends: Evidence From Taiwan. J Med Internet Res 2015;17(11):e264 [FREE Full text] [doi: 10.2196/jmir.4516] [Medline: 26586281]

19. Ling R, Lee J. Disease Monitoring and Health Campaign Evaluation Using Google Search Activities for HIV and AIDS, Stroke, Colorectal Cancer, and Marijuana Use in Canada: A Retrospective Observational Study. JMIR Public Health Surveill 2016 Oct 12;2(2):e156 [FREE Full text] [doi: 10.2196/publichealth.6504] [Medline: 27733330]

20. Ayers JW, Althouse BM, Allem J, Rosenquist JN, Ford DE. Seasonality in seeking mental health information on Google. Am J Prev Med 2013 May;44(5):520-525. [doi: 10.1016/j.amepre.2013.01.012] [Medline: 23597817]

21. Fahy E, Hardikar R, Fox A, Mackay S. Quality of patient health information on the Internet: reviewing a complex and evolving landscape. Australas Med J 2014;7(1):24-28 [FREE Full text] [doi: 10.4066/AMJ.2014.1900] [Medline: 24567763]

22. Phillips CA, Barz LA, Li Y, Schapira MM, Bailey LC, Merchant RM. Relationship Between State-Level Google Online Search Volume and Cancer Incidence in the United States: Retrospective Study. J Med Internet Res 2018 Jan 08;20(1):e6 [FREE Full text] [doi: 10.2196/jmir.8870] [Medline: 29311051]

23. Xu C, Yang H, Sun L, Cao X, Hou Y, Cai Q, et al. Detecting Lung Cancer Trends by Leveraging Real-World and Internet-Based Data: Infodemiology Study. J Med Internet Res 2020 Mar 12;22(3):e16184 [FREE Full text] [doi: 10.2196/16184] [Medline: 32163035]

24. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]

25. Ruths D, Pfeffer J. Social sciences. Social media for large studies of behavior. Science 2014 Nov 28;346(6213):1063-1064. [doi: 10.1126/science.346.6213.1063] [Medline: 25430759]

26. Reyes G. Leveraging behavioral insights in the age of big data. World Bank Blogs. 2017. URL: https://blogs.worldbank.org/latinamerica/leveraging-behavioral-insights-age-big-data?cid=SHR_BlogSiteShare_EN_EXT [accessed 2020-10-31]

27. Afif Z, Islan W, Calvo-Gonzalez O. Behavioral science around the world: Profiles of 10 countries. World Bank. 2018. URL: https://documents1.worldbank.org/curated/en/710771543609067500/pdf/132610-REVISED-00-COUNTRY-PROFILES-dig.pdf [accessed 2020-10-31]

28. Butler D. When Google got flu wrong. Nature 2013 Feb 14;494(7436):155-156. [doi: 10.1038/494155a] [Medline: 23407515]

29. Mills KA. What are the threats and potentials of big data for qualitative research? Qualitative Research 2017 Nov 30;18(6):591-603. [doi: 10.1177/1468794117743465]

30. Compare 'aids' and 'hiv'. Google Trends. URL: https://trends.google.com/trends/explore?date=2004-01-01%202019-12-31&geo=ZA&q=aids,hiv [accessed 2020-10-31]

31. Hipel K, McLeod A. Time Series Modelling of Water Resources and Environmental Systems, 1st edition. New York: Elsevier Science; 1994.

32. Hollander M, Wolfe D, Chicken E. Nonparametric statistical methods, 2nd edition. Hoboken, New Jersey: John Wiley & Sons; 2013.

33. Wikipedia Afrikaans Pageviews Analysis on HIV and AIDS 2015-2020. Wikipedia. URL: https://pageviews.toolforge.org/?project=af.wikipedia.org&platform=all-access&agent=user&redirects=0&start=2015-07&end=2020-10&pages=HIV|AIDS [accessed 2020-11-05]

34. HIV estimates with uncertainty bounds 1990-Present. UNAIDS. URL: https://www.unaids.org/en/resources/documents/2020/HIV_estimates_with_uncertainty_bounds_1990-present [accessed 2020-10-31]

35. Mavragani A, Ochoa G, Tsagarakis KP. Assessing the Methods, Tools, and Statistical Approaches in Google Trends Research: Systematic Review. J Med Internet Res 2018 Nov 06;20(11):e270 [FREE Full text] [doi: 10.2196/jmir.9366] [Medline: 30401664]

36. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]

37. Syamsuddin M, Fakhruddin M, Sahetapy-Engel JTM, Soewono E. Causality Analysis of Google Trends and Dengue Incidence in Bandung, Indonesia With Linkage of Digital Data Modeling: Longitudinal Observational Study. J Med Internet Res 2020 Jul 24;22(7):e17633 [FREE Full text] [doi: 10.2196/17633] [Medline: 32706682]

38. Yuan K, Huang G, Wang L, Wang T, Liu W, Jiang H, et al. Predicting Norovirus in the United States Using Google Trends: Infodemiology Study. J Med Internet Res 2021 Sep 29;23(9):e24554 [FREE Full text] [doi: 10.2196/24554] [Medline: 34586079]

39. An L, Russell DM, Mihalcea R, Bacon E, Huffman S, Resnicow K. Online Search Behavior Related to COVID-19 Vaccines: Infodemiology Study. JMIR Infodemiology 2021 Nov 12;1(1):e32127 [FREE Full text] [doi: 10.2196/32127] [Medline: 34841200]

40. Johnson AK, Bhaumik R, Tabidze I, Mehta SD. Nowcasting Sexually Transmitted Infections in Chicago: Predictive Modeling and Evaluation Study Using Google Trends. JMIR Public Health Surveill 2020 Nov 05;6(4):e20588 [FREE Full text] [doi: 10.2196/20588] [Medline: 33151162]

41. Chiu APY, Lin Q, He D. News trends and web search query of HIV/AIDS in Hong Kong. PLoS One 2017 Sep 18;12(9):e0185004 [FREE Full text] [doi: 10.1371/journal.pone.0185004] [Medline: 28922376]

42. Mavragani A, Ochoa G. Forecasting AIDS prevalence in the United States using online search traffic data. J Big Data 2018 May 19;5(1):1-21. [doi: 10.1186/s40537-018-0126-7]

43. Al-Rawi AK. Exploring HIV-AIDS interests in the MENA region using Internet based searches. First Monday. URL: http://hdl.handle.net/1765/102387 [accessed 2020-02-23]

44. Chrzanowski J, Sołek J, Fendler W, Jemielniak D. Assessing Public Interest Based on Wikipedia's Most Visited Medical Articles During the SARS-CoV-2 Outbreak: Search Trends Analysis. J Med Internet Res 2021 Apr 12;23(4):e26331 [FREE Full text] [doi: 10.2196/26331] [Medline: 33667176]

45. Seo D, Jo M, Sohn CH, Shin S, Lee J, Yu M, et al. Cumulative query method for influenza surveillance using search engine data. J Med Internet Res 2014 Dec 16;16(12):e289 [FREE Full text] [doi: 10.2196/jmir.3680] [Medline: 25517353]

46. Zhang D, Shi Z, Hu H, Han GK. Classification of the Use of Online Health Information Channels and Variation in Motivations for Channel Selection: Cross-sectional Survey. J Med Internet Res 2021 Mar 09;23(3):e24945 [FREE Full text] [doi: 10.2196/24945] [Medline: 33687342]

XSL•FO

RenderX

Original Paper

Performance of an eHealth (NOMHAD) System Comprising Telemonitoring, Telenotification, and Telecoaching for Patients With Multimorbidity: Proof-of-Concept Study

Mathieu Chacornac[1], MD; Aurélien Faoro, MD; Joëlle Texereau[2,3], MD, PhD; Catherine Billoet[4], MD; Stéphane Hominal[5], MD

[1]Department of Cardiology, Centre Hospitalier Annecy-Genevois, Pringy, France

[2]Department of Respiratory Medicine, Cochin Hospital, Assistance Publique-Hôpitaux de Paris, Université de Paris, Paris, France

[3]VitalAire, Air Liquide Health Care, Gentilly, France

[4]Air Liquide Santé International, Paris, France

[5]Department of Pneumology, Centre Hospitalier Annecy-Genevois, Pringy, France

Corresponding Author:
Mathieu Chacornac, MD
Department of Cardiology
Centre Hospitalier Annecy-Genevois
1, avenue de l'Hôpital
Pringy, 74374
France
Phone: 33 450636749
Email: mchacornac@ch-annecygenevois.fr

Abstract

Background: Management of patients with multiple chronic diseases is a growing public health challenge, especially in rural sectors where access to physicians may be limited. Connected medical devices monitoring vital signs, associated with eHealth program and structured telephone support, may improve complex patient management through early detection of disease complications, positive impact on patients' health, and health resources consumption optimization.

Objective: The aim of this study was to evaluate the technical performance and user experience of the NOMHAD eHealth system in patients with multimorbidity.

Methods: This was a pilot, single-arm, interventional study. Patients with multimorbidity with any combination of chronic heart failure (CHF), chronic obstructive pulmonary disease, and diabetes were followed for 80-100 days using the NOMHAD eHealth system. This system used connected devices telemonitoring symptoms and vital signs (eg, body weight, oxygen saturation, pulse rate, blood pressure, and blood glucose), associated with structured telecoaching and educational support by call center nurses. An overall risk indicator (ORI) was automatically computed after each data teletransmission. The ORI was color coded; green indicated no action required; yellow, orange, and red (low to high priority, respectively) generated telenotifications and indicated to the nurses the need for a telecoaching action. Each ORI was calculated by combining 7 clinical stability system indicators based on symptom questionnaires and vital signs. Technical accuracy of the system was assessed by comparing system-generated ORIs with ORIs recalculated from raw data. Ease of use, usefulness, satisfaction, and acceptability of the system were assessed through patient adherence to self-assessments, and through self-administered questionnaires to patients, call center nurses, and physicians.

Results: A total of 23 patients were enrolled in this study and participated between April 2016 and March 2017 at 5 study centers in France. All patients were successfully equipped and evaluable for analysis. Mean age was 68.5 (SD 10.4) years and most patients were men (n=20). The most common multimorbidity was CHF + diabetes (n=15), followed by patients with all 3 diseases (n=5). Mean effective follow-up was 78.7 (SD 24.2) days. The system generated 6263 ORIs, as several ORIs could be generated on a single day for any patient. Overall system sensitivity was 99.2% (95% CI 98.9-99.4) and overall specificity was 91.3% (95% CI 87.7-94.1). Most patients (20/23, 87%) were satisfied with the system and agreed that it helped them to better understand and manage their diseases, and 19/23 (83%) valued the nurse regular contacts. Nurses and physicians were generally satisfied with the system and considered it useful. All users indicated they would agree to long-term use of the system.

Conclusions: This study provides evidence that the NOMHAD eHealth system is accurate, acceptable, informative, and feasible for patients with multimorbidity, supporting further investigation of its clinical benefits.

Introduction

Background

The management of patients with chronic diseases is a major public health challenge that continues to grow as the numbers of these patients continue to increase [1]. In developed countries, advances in medical care and consequently longer life expectancies have contributed to a steady rise in the numbers of patients with chronic diseases such as chronic heart failure (CHF), chronic obstructive pulmonary disease (COPD), and diabetes, all of which are among the leading causes of death [2]. These patients now account for a major share of many national health care budgets. In France, it was estimated that approximately 80% of the national health care budget in 2012 was dedicated to care for patients with chronic illnesses [3].

Medical deserts due to uneven geographic distribution of health care practitioners are a reality for many patients in France. Despite a steady increase in the number of physicians, many will retire in the next few years, which will further decrease health care availability in some territories. This trend is exacerbated by a large disparity across France in the numbers of physicians available per person, which is highest near Paris [4].

Faced with the rising prevalence and burden of chronic illnesses, health care systems have recognized that traditional reactive medical care involving repeated hospitalizations for disease complications is unsustainable [5,6]. Thus, alternative remote methods of managing patients with chronic diseases are needed to promote and stabilize patient health at home, while minimizing the risk of hospitalization and the burden on health care resources. Chronic disease telemanagement programs aim to foster continuity between hospital and home care, prevent disease worsening and the onset of complications, prevent or reduce hospitalizations, ensure appropriate and timely treatment when needed, educate patients to empower disease self-management, and optimize physician involvement. The ultimate goal for patients is to live safely at home while continuing to receive optimal care.

French national recommendations have called for the development of such telemedicine interventions in patients with CHF, COPD, and diabetes—the 3 chronic diseases that impose the highest burdens on the health care system. The sudden episodes of deterioration that characterize these diseases, such as cardiac decompensations in CHF, acute exacerbations in COPD, and major hypoglycemic events in diabetes, require hospitalization and very often a second hospitalization soon after discharge. Some of these hospitalizations and

rehospitalizations may be partially preventable if symptom worsening is detected and managed early [7,8].

Home telemonitoring programs are among the solutions currently investigated for remote disease management. The intent is to provide high-quality health care to patients while they live at home. In patients with CHF, remote telemonitoring programs have been shown to reduce mortality [9-13], hospitalizations [9,10,12-14], emergency care unit visits [9], and health care costs [9]. In a review of transitional care interventions, programs that included home visits and multidisciplinary interventions reduced mortality and all-cause readmissions, but telemonitoring had no beneficial effect [15]. In patients with COPD, the benefits of telemonitoring programs versus usual care have been less clear. In the recent COPD Patient Management European Trial (COMET) that investigated a COPD disease management intervention that included remote telemonitoring in patients with severe COPD, acute care hospitalization days and mortality were significantly lower in the disease management group [16]. Other analyses have found inconsistent impact of telemonitoring on hospitalizations [17-19] and quality of life [17-20], a trend toward reduced costs [20,21], improved physical activity but no effect on physical capacity or dyspnea [22], and no effect on length of hospital stay or mortality [19]. Most reviewers concluded that despite some encouraging results, additional evidence from larger, high-quality studies is needed before definitive conclusions and recommendations can be made on the benefits of telemonitoring programs for patients with COPD [12,19,20,22,23]. For patients with diabetes, clinical trials and systematic reviews have consistently demonstrated that telemonitoring programs are feasible and can sustain improved glycemic control over time versus usual care [24-26]. Evidence for other benefits, such as reduced BMI [24], reduced use of health care resources and costs [26,27], and lower mortality and cognitive decline in older patients, has been modest or inconsistent [27].

Some of the difficulty in interpreting the results of the many studies investigating the benefits of telemonitoring programs in patients with chronic diseases stems from the heterogeneity in the programs and in the patient populations studied. Most telehealth programs have focused on a single chronic disease, but the reality is that many patients with chronic diseases have more than 1 disease [28-30]. The overall prevalence of such multimorbidity was 23% among all patients in a national health care database in Scotland [28], whereas the prevalence of multimorbidity is much higher (>60%) in the growing European population of patients over the age of 65, who tend to accumulate chronic conditions as they age [28,30]. Multimorbidity is well described in patients with CHF, COPD,

XSL•FO

RenderX

or diabetes. In another study conducted in Scotland, 42% of patients with CHF also had diabetes and 26% also had COPD [27]. Similarly, 39% of patients with COPD had CHF and 22% had diabetes, and 45% of patients with diabetes had CHF and 15% had COPD [27]. Thus, telehealth and telemonitoring programs designed for use in patients with a single chronic disease may be limited in their capacity to provide benefits to patients with multiple chronic conditions or comorbidities.

The NOMHAD eHealth System: Overview

The NOMHAD eHealth system, consisting of a web interface for physicians and nurses (NOMHAD Chronic) and an app for patients (NOMHAD Mobile), is derived from a Spanish system and is specifically designed to handle patients with multiple chronic diseases [31]. This system aims to provide telemonitoring by means of detecting health status deterioration based on connected devices and telecoaching based on educational support by dedicated call center nurses.

Study Objectives

In this exploratory study, the technical performance of the NOMHAD eHealth system was evaluated in patients with multimorbidity with any combination of CHF, COPD, and diabetes. The ease of use, usefulness, satisfaction, and acceptability of the system were also investigated from the perspectives of the patients, call center nurses, and physicians.

Methods

Study Design and Objectives

This was a pilot, multicenter, nonrandomized, single-arm, open-label, uncontrolled interventional study to evaluate the performance of the NOMHAD eHealth system of telemonitoring, telenotification, and telecoaching (Agence Nationale de Sécurité du Médicament et des Produits de Santé [ANSM] registration number [ID-RCB]: 2015-A01106-43). The system was evaluated over 3 months of follow-up in a single group of approximately 30 ambulatory patients with multimorbidity.

The primary objective of the study was to evaluate the technical performance of the NOMHAD eHealth system in this population. This objective was to be achieved by comparing patient overall risk indicators (ORIs) generated by the system

with those recalculated on the basis of raw data teletransmitted to the software platform. Secondary objectives were to evaluate the impressions and perspectives of NOMHAD eHealth system users (patients, study physicians, and nurses) using different self-questionnaires about overall ease of use, usefulness, satisfaction, and acceptability. Feasibility of the intervention, technological performance of the system, clinical events, adverse events (AEs), and adverse device events were also assessed.

Ethics Approval

The study protocol (2015-A01106-43) and subsequent amendments were approved by a local independent ethics committee (Comité de Protection des Personnes Lyon Sud-Est IV) and the French Medicine Agency (ANSM). The study was conducted in compliance with Good Clinical Practice guidelines, the Declaration of Helsinki, the EU Council Directive 93/42/EEC regarding Medical Devices, and International Standard ISO 14155. Study protocol, benefits, and risks were explained to participating patients, who were required to provide written informed consent.

Study Participants

Patients were recruited and enrolled at 5 study centers in France (Rhône-Alpes region). Adults over the age of 18 years with at least two chronic diseases among CHF, COPD, and diabetes and with at least one hospitalization for acute decompensation of CHF or exacerbation of COPD during the year prior to inclusion were eligible to participate in the study. Patients were also required to speak and understand French and have a landline phone. To be enrolled, patients had to meet the French Health Authorities (Haute Autorité de Santé) diagnosis criteria [32], briefly summarized in Table 1. No biological or clinical examinations were required at study entry, but the results of the most recent examinations performed per routine practice were to be reported. Patients were excluded if they were pregnant or breastfeeding, were institutionalized, had a life expectancy of less than 3 months, had received within the past 6 months or were receiving chemotherapy or radiotherapy for cancer, required dialysis for chronic renal insufficiency, had a condition likely to interfere with study procedures, were not covered by the French National Health System, or had participated within the past 30 days or were participating in another interventional trial.

Table 1. Main diagnosis criteria[a] for enrollment of patients with chronic obstructive pulmonary disease, diabetes, or chronic heart failure.

Disease	Criteria
Chronic obstructive pulmonary disease	Clinical signs AND exposure to a known risk factor AND forced expiratory volume in 1 second/forced vital capacity ratio <70% after bronchodilator
Diabetes	Blood sugar >1.26 g/L (7.0 mmol/L) after 8-hour fasting (assessed twice) OR diabetes symptoms with blood sugar >2 g/L (11.1 mmol/L) OR blood sugar >2 g/L, 2 hours after a 75-g oral glucose challenge
Chronic heart failure	Clinical symptoms or signs AND objective evidence of structural or functional cardiac abnormality at rest. Systolic heart failure OR heart failure with preserved ejection fraction

[a]Criteria from the French National Health Authorities (Haute Autorité de Santé [32]).

NOMHAD eHealth System

The NOMHAD eHealth system employs the NOMHAD Chronic software platform (version 1.6.4; Connected Health Services)

and the remote NOMHAD Mobile app (version 1.2.1) installed on an interactive tablet (Figure 1). Through NOMHAD Mobile, patients self-measure their vital signs through connected devices

XSL·FO
RenderX

(or manually), provide additional health status–related information by answering questionnaires, and transmit the information to the central database.

The NOMHAD Chronic software platform consists of a web interface for physicians to define personalized patient care plans and browse patient data, and a call center station web interface for nurses (skilled home health care provider nurses working at the VitalAire France nurse call center) to browse patient data, prioritize nurse activities, and report telecoaching actions.

Figure 1. Overview of the NOMHAD eHealth system. The NOMHAD eHealth system consists of the NOMHAD Chronic software platform and the remote NOMHAD Mobile app installed on an interactive tablet, which is used by the patient to enter vital sign data, either manually or through connected devices, and answer questionnaires. These data are transferred to the NOMHAD Chronic software platform using available 3G, 4G, or Wi-Fi connections. The NOMHAD eHealth system is a global system including health care technicians, call center nurses, study physicians, and referring physicians. 3G: third generation; 4G: fourth generation; SpO$_2$: arterial oxygen saturation; Wi-Fi: wireless fidelity.
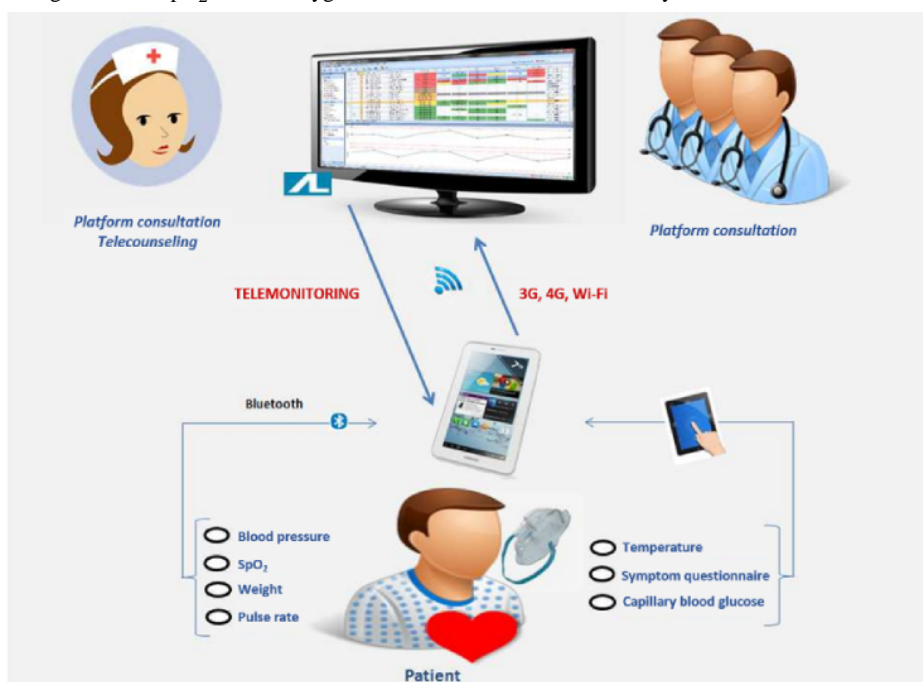
NOMHAD Chronic monitors up to 7 clinical stability systems indicators that inform about chronic disease/health status parameters according to patient medical conditions and individual-specific thresholds: hemodynamic stability, respiratory stability, pain/sleep quality, diet/diuresis, tissue integrity, cognitive system/mood, and medication.

The software platform incorporates an algorithm that calculates the ORI to guide the nurse in prioritizing telecoaching actions. ORIs are color coded, enabling immediate identification of the patients to be contacted and their level of priority. A green ORI indicates that no action is required, whereas yellow (low priority), orange (medium priority), and red (high priority) ORIs generate telenotifications and indicate to the call center nurse the need for a telecoaching action, from simple reinforcement, repeating measurements and adapting the care plan, to advising the patient to call an emergency medical service. The ORI is calculated by combining the 7 clinical stability system indicators, which are themselves calculated by combining the results of symptom questionnaires and vital signs. An individual ORI is calculated by the algorithm every time a new value of any monitored parameter is transmitted by the patient. Therefore, several ORIs and corresponding telenotifications may be generated on a single day for a given patient.

Patient Interventions

Patients participated in the study over a period of approximately 3 months (Figure 2). Inclusion was followed by an adaptation period during which the telemonitoring kit was installed by a home health care provider technician and first used autonomously by the patients in their home. The follow-up period began with a phone interview with the nurse to evaluate the patient's self-management skills specific for their chronic diseases, establish objectives for improvement, and collect patient lifestyle data. During the follow-up period, patients performed regular self-assessments of their health status according to their chronic diseases. Patients used the tablet to enter their answers to health questionnaires and the devices connected to the tablet to measure vital signs (eg, blood pressure, pulse rate, arterial oxygen saturation, daily blood glucose, body weight depending on their medical condition, and personalized care plan) and transmit them to the NOMHAD Chronic system. The standard patient care plan (consisting of self-measurements twice a week and questionnaires once a week in stable state) could be personalized according to patient needs, the system notifications generated, and study physician assessment.

Figure 2. Study flow for the patient. The total study duration for a patient was approximately 15 weeks.
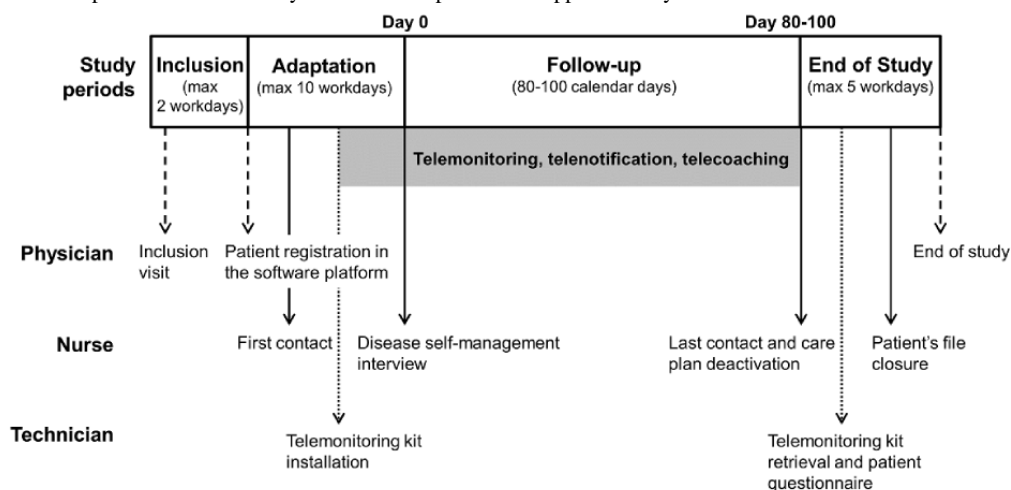
Patients were contacted by a nurse at regular intervals (at least every 15 days for the first 2 months) and when the system generated a telenotification. Patients could also contact a nurse whenever necessary during regular business hours via a toll-free phone number, for instance, to report a hospitalization. Through the NOMHAD Mobile app, patients could consult disease self-management support modules installed on their tablet and could voluntarily participate with other study patients in online video classroom sessions conducted by the nurses. The objective of these sessions was to provide patients with supplementary information about their chronic diseases and their management while taking advantage of group dynamics. The telemonitoring program could be suspended temporarily by a call center nurse for patient's personal reasons or during a hospitalization.

Nurses from the call center were trained to provide support by phone to patients with chronic diseases and were connected to the NOMHAD Chronic system to manage telenotifications (according to their priority level) during regular working hours. They performed standardized telecoaching actions, such as advising patients to seek physician consultation or emergency care, according to the type and severity of the telenotifications.

Physicians could access patient data at any time; were prompted to log-on to the system during patient consultations; and could modify patient medical data, medications, procedures, system threshold values for health status parameters, and recorded any AE. After a follow-up period of 80-100 days, the care plan was deactivated, equipment was retrieved, and study physicians completed the electronic case report form.

Outcomes and Evaluation Methods

For the primary performance criterion, the sensitivity and specificity of the ORIs generated during the follow-up period were assessed by comparing the ORIs generated by the NOMHAD system using a specific algorithm (ORI_{NOMHAD}) with those expected when recalculated from the raw data teletransmitted to the software platform (ORI_{EXP}) based on its algorithm specifications. Each combination of ORIs (ORI_{NOMHAD} vs ORI_{EXP}) obtained was categorized as follows: if the ORI_{NOMHAD} and ORI_{EXP} were both green, the scoring was considered "correctly not telenotified" (ie, true negative [TN]);

if the ORI_{NOMHAD} and ORI_{EXP} colors were the same for yellow, orange, or red, the scoring was considered "correctly telenotified" (ie, true positive [TP]); if the ORI_{NOMHAD} priority color was lower than the ORI_{EXP} color, the scoring was considered "incorrectly identified as lower priority" (ie, false negative [FN]); and if the ORI_{NOMHAD} priority color was higher than the ORI_{EXP} color, the scoring was considered "incorrectly identified as higher priority" (ie, false positive [FP]). Secondary performance criteria were the numbers and frequencies of telenotifications and the sensitivity (defined as $TP/[TP + FN] \times 100$) and specificity (defined as $TN/[TN + FP] \times 100$) of the clinical stability system indicators. The overall sensitivity and specificity were calculated from all ORI recordings.

At the end of the study, patients, nurses, and physicians completed user-specific self-questionnaires to assess their perspectives on the ease of use, usefulness, and satisfaction of the system. These self-questionnaires were developed by the sponsor specifically for the study. System acceptability was assessed from the actual duration of follow-up and the duration of effective follow-up (after deduction of the temporary suspensions periods). System feasibility was assessed from the numbers of patients refusing or discontinuing the program for any nonmedical reasons. Safety was evaluated from device deficiencies and AEs starting during the study follow-up period.

Sample Size Estimation

An appropriate sample size could not be calculated because no assumptions could be made prior to the start of the trial regarding the primary technical performance criterion in patients with multimorbidity. It was estimated that approximately 30 evaluable patients would be sufficient to establish preliminary descriptive conclusions for the primary criterion.

Statistical Methods

Statistical analyses were performed using SAS software, version 9.2 (SAS Institute). Summary statistics for continuous variables were number of observations, mean, SD, median, first and third quartiles (Q1 and Q3), minimum and maximum values, and the number of missing information. Summary statistics for dichotomous or categorical variables were the number and percentage of each of the scores or categories with 95% CIs

calculated using the exact binomial distribution for percentages, where applicable. The agreement between all pairs of ORIs generated by the software platform and by those recalculated on the basis of the raw data teletransmitted (each time the algorithm was run by the software platform) was evaluated with SAS by calculating the weighted Fleiss–Cohen κ coefficient for all pairs of ORI scores and by testing the null hypothesis that this coefficient was equal to 0 (ie, no agreement apart from pure chance agreement existed between both scorings). Results are presented for all patients who were enrolled in the study.

Amendments to the Study Protocol

After the study began, the study protocol was amended to stop recruitment before the planned number of enrolled patients was reached. This was done because the recruitment period had already been extended and it revealed that investigators had reached their maximal recruitment potential based on the study inclusion criteria.

Results

Study Patients

A total of 23 patients were enrolled in the study and participated between April 2016 and March 2017 at 5 study centers. Three patients did not complete the study: 1 withdrew consent, 1 moved into a care facility that precluded study procedures, and 1 requested to discontinue telemonitoring. Thus, 20 patients completed study over a mean duration of 99.7 (SD 18.9) days. There were 18 patients with at least one minor protocol deviation, most were for missing data (n=15) or time window deviations (n=8). No major protocol deviations (defined as those affecting the minimal 2-month required self-recordings) occurred, except the 7-day follow-up of the patient who withdrew consent.

The mean age of the patients in the study was 68.5 (SD 10.4) years and most patients (20/23, 87%) were men (Table 2). The mean BMI was 33.9 (SD 7.2) kg/m^2 with a range of 24.5-54.6 kg/m^2. Most patients did not live alone, had a caregiver, and declared having an occupational activity outside the home and regular physical activity.

Table 2. Patient demographics, lifestyle, therapy, and vital signs at study entry.

Characteristic	All patients (N=23)
Sex, male, n (%)	20 (87)
Age (years), mean (SD)	68.5 (10.4)
BMI (kg/m^2), mean (SD)	33.9 (7.2)
Never or former smoker, n (%)	21 (95)[a]
Living alone[b], n (%)	8 (35)
Presence of a caregiver[b], n (%)	13 (57)
Occupational activity outside the house[b], n (%)	14 (61)
Regular physical activity[b], n (%)	15 (65)
Difficulties in understanding the diet[b], n (%)	11 (48)
Regular body weight monitoring[b], n (%)	17 (74)
Blood glucose self-monitoring (patients with diabetes)[b], n (%)	18 (86)[c]
Therapy[d], n (%)	
Insulin	11 (48)
Home oxygen therapy	3 (13)
Home noninvasive ventilation	5 (22)
Systolic blood pressure (mmHg), mean (SD)	125.4 (23.3)
Diastolic blood pressure (mmHg), mean (SD)	72.8 (13.6)
Pulse rate (beats/minute), mean (SD)	79.7 (19.0)

[a]N=22.

[b]As declared by the patient to the call center nurse.

[c]N=21.

[d]As recorded in the NOMHAD eHealth system. Multiple responses were allowed. % = (n row/n nonmissing) × 100, except for multiple responses where % = (n row/N group) × 100.

XSL•FO
RenderX

The most frequent multimorbidity was CHF + diabetes (n=15; Table 3). Five patients had all 3 diseases (CHF + diabetes + COPD). The severity of CHF was most frequently reported as New York Heart Association class II (11/19, 58%) or III (7/19, 37%), and patients with CHF (n=21) had a mean left ventricular ejection fraction of 44.5% (SD 13.5%). Patients with COPD (n=6) had a mean forced expiratory volume in 1 second of 55.0% (SD 16.9%) predicted (n=6). Patients with diabetes had

mainly type 2 diabetes (n=19), with a mean last measured glycosylated hemoglobin (HbA_{1c}) of 7.7% (SD 1.6%) (n=18). Micro- and macro-angiopathic complications were reported in 11 and 4 patients, respectively. Overall, patients had a mean of 6.5 (SD 2.9) concomitant diseases that were recorded by their study physician at inclusion in addition to their targeted chronic diseases.

Table 3. Patient multimorbidity.

Multimorbidity profile of study patients	All patients (N=23)
Patient's main chronic disease profile, n (%)	
CHF[a] + COPD[b]	2 (9)
CHF + diabetes	15 (65)
COPD + diabetes	1 (4)
CHF + COPD + diabetes	5 (22)
Number of concomitant diseases at inclusion, mean (SD)	6.5 (2.9)
Comorbidities[c], n (%)	
Dyslipidemia	16 (70)
Arterial hypertension	16 (70)
Chronic respiratory failure	7 (30)
No comorbidity ticked	3 (13)

[a]CHF: chronic heart failure.

[b]COPD: chronic obstructive pulmonary disease.

[c]As recorded in the NOMHAD eHealth system. Multiple responses were allowed. % = (n row/n nonmissing) × 100, except for multiple responses where % = (n row/N group) × 100.

System Performance

For all patients, the NOMHAD eHealth system generated a total of 6263 ORIs that could be recalculated from the raw data teletransmitted during the 3-month follow-up period. Among these ORIs, 294 (4.69%) were green, 4480 (71.53%) were yellow, 1441 (23.01%) were orange, and 48 (0.77%) were red. ORI proportions by color were similar for each chronic disease individually.

For all patients, there was good agreement between the telenotifications generated by the NOMHAD eHealth system and those recalculated from the raw data. In the primary analysis, overall sensitivity was 99.2% (95% CI 98.9-99.4) and overall specificity was 91.3% (95% CI 87.7-94.1). Sensitivity and specificity results were similar for each chronic disease individually. Agreement evaluation of all ORI pairings yielded a weighted Fleiss–Cohen κ coefficient of 0.975 (95% CI 0.969-0.981; $P<.001$), indicating good agreement between all pairs that was not due to pure chance. Kappa coefficients for each chronic disease individually were similarly high (≥0.975).

System Effectiveness

During the 3-month follow-up period, all patients generated at least one telenotification and triggered, therefore, nurse telecoaching actions. The highest proportions of telenotifications were yellow and orange (ie, low and moderate priority, respectively). All 23 patients generated at least one yellow

ORI-low priority telenotification (with a mean of 190 per patient over the 3-month period), 17 patients generated at least one orange ORI-moderate priority telenotification (with a mean of 83 per patient), and 7 patients generated at least one red ORI-high priority telenotification (with a mean of 6 per patient).

The management of the telenotifications resulted in a median of 21 nurse contacts per patient (Q1-Q3, 6-49). The call duration lasted 11-20 minutes at least once for most patients (22/23, 96%). The mean time interval between a telenotification during business hours and registration of telecoaching action was 68.3 (SD 63.6) minutes.

Physicians did not change the predefined parameters or thresholds. On average, physicians connected 3 times per patient for approximately 30 minutes in total, during the 3-month follow-up.

End User Perspectives

Patients

Most (21/23) patients felt that health care technicians spent sufficient time installing the study devices and training them to use the eHealth system. Most patients were satisfied with the coaching they received from the nurses.

Eighteen patients agreed or strongly agreed that the information contained in the app was useful in helping them better understand their illnesses and 14 patients agreed or strongly agreed that it helped them better manage their symptoms. Patient

contact with the call center nurses was considered helpful, as 16 patients agreed or strongly agreed that it helped them become more independent in managing their illnesses, and 17 patients agreed or strongly agreed that they received effective advice when needed. Sixteen patients participated in at least one online video classroom session. However, only 7 patients agreed or strongly agreed that the conferences helped them to better cope with their illnesses through sharing their experience with other patients.

Overall, 20 patients agreed or strongly agreed that they were generally satisfied with the eHealth system, 10 patients agreed or strongly agreed that it improved their relationship with their study physician, and 15 patients agreed or strongly agreed that they would be prepared to use it long term.

Nurses

Nurses (n=3) strongly agreed that their involvement in the study helped them to communicate more effectively with patients and agreed that it helped patients to better understand how to manage their diseases with self-measurement and self-management tools. All nurses strongly agreed that they would be prepared to use the system long term. Overall, they ranked the remote vital signs measurements and the planned regular phone calls with the patients to be the best components.

Study Physicians

Study physicians (n=5) agreed that the software platform was somewhat or fairly intuitive and user-friendly, and 4 agreed that its content met their expectations. One of them modified the standard monitoring patient care plan of 1 or more patients and felt this was a medically relevant action and easy to perform.

Three physicians agreed that the telemonitoring and telecoaching interventions by the nurses resulted in physicians seeing only those patients with a genuine need, whereas all agreed that the frequency of notifications and patient consultations was sufficient.

Three physicians agreed or strongly agreed that the information available on the software platform enabled them to better understand their patients' disorders and to better manage their patients. Four of the physicians agreed or strongly agreed that they could better anticipate complications affecting patients with multiple morbidities and that their patients were better equipped with self-measurement and self-management tools. Four agreed or strongly agreed that it improved their relationship with their patients and 4 were fairly or very prepared to use the system long term. The study physicians ranked the vital signs, the availability of the telemonitoring platform, and the telecoaching actions among the more preferred components.

System Acceptability, Feasibility, and Technical Performance

Acceptability, as assessed by the extent to which planned events were completed, was generally high (Table 4). An average of 87.3% (SD 34.6) of the total planned regular phone contacts every 2 weeks were performed by the patients. Acceptability was lower for the online video classroom sessions. Out of a total of 8 different contents, patients actively participated in an average of 4.7 (SD 5.8) sessions. Patients consulted the chronic disease self-management support modules available on their tablet an average of 11.2 (SD 15.2) times during the 3-month follow-up and spent an average total time of approximately 16 hours on these consultations (median 5 hours).

Table 4. Acceptability of the nomhad eHealth system by the patients.

Variable	All patients (n=23)
Follow-up	
Duration of follow-up period (days), mean (SD)	85.8 (19.0)
Duration of effective follow-up period (days), mean (SD)	78.7 (24.2)
Duration of system temporary suspension (days), mean (SD)	7.2 (14.4)
Ratio of duration of effective follow-up to duration of follow-up period (%), mean (SD)	90.1 (18.2)
Planned regular phone contacts with the call center nurse	
Number of actual planned phone contacts, mean (SD)	3.5 (1.4)
Number of planned phone contacts scheduled, mean (SD)	3.7 (0.8)
Ratio of actual to planned number of phone contacts (%), mean (SD)	87.3 (34.6)
Online video classroom sessions	
Number of attended sessions (any type), mean (SD)	4.7 (5.8)
Total duration of actual participation (minutes), mean (SD)	136.3 (206.7)
Ratio of duration of actual participation to overall session duration (%), mean (SD)	60.5 (39.6)
Information modules	
Number of consultations, mean (SD)	11.2 (15.2)
Total duration of consultations (minutes), median (Q1-Q3)	310.4 (3-1561)

The program also appeared to be feasible. No eligible patients refused to participate in the study. One patient withdrew consent after 1 week of follow-up but no patients discontinued prematurely due to lack of connection or inability to use the

XSL•FO
RenderX

study equipment. Eight patients (35%) required at least one maintenance visit and 6 patients (26%) needed at least one device replacement or had technical issues with the study equipment. Data transmission failures were mostly due to connection issues related to the mountainous study area, and represented an average of 3.4% (SD 5.9%) of the total number of data transmissions per patient.

Safety Evaluations

Nine patients (39%) experienced a total of 16 AEs during their follow-up period. None of these AEs were considered to be related to the study medical device or procedures and no study discontinuation resulted from any AE. Seven patients (30%) experienced a total of nine serious AEs (Table 5), all of which were related to unscheduled hospitalizations. All these serious AEs were considered by the investigators to be related to the chronic diseases under investigation or to other diseases. The most frequent serious AEs were cardiac failures (3/23, 13%).

Table 5. Serious adverse events occurring during study follow-up.

Sex/age	Chronic diseases	Serious adverse event (MedDRA-preferred term)[a]	Time from inclusion to onset (days)	Duration (days)	Severity	Time since last ORI[b] (days)	Last color of ORI[c,d]	Last color of SSI[e,f]
Male/82 years	CHF[g]/diabetes	Cardiac failure	32	4	Moderate	2	Yellow	• SSIH[h]: Blue • SSIR[i]: Green • SSIP[j]: Green • SSID[k]: Green • SSIT[l]: Green • SSIM[m]: Green
Male/62 years	CHF/diabetes	Cardiac failure/erysipelas	52	20	Severe	24[n]	Yellow	• SSIH: Green • SSIR: Green • SSIP: —[o] • SSID: Green • SSIT: — • SSIM: Green
Male/72 years	CHF/diabetes	Atrial fibrillation	38	8	Moderate	2	Yellow	• SSIH: Green • SSIR: Blue • SSIP: Green • SSID: Orange • SSIT: Green • SSIM: Green
Male/84 years	CHF/COPD/diabetes	Cardiac failure	38	14	Severe	1	Yellow	• SSIH: Blue • SSIR: Green • SSIP: Green • SSID: Orange • SSIT: Green • SSIM: Green
Female/66 years	CHF/COPD/diabetes	COPD/alveolitis	86	11	Moderate	9	Yellow	• SSIH: Green • SSIR: Green • SSIP: Green • SSID: Green • SSIT: — • SSIM: Green
Male/62 years	CHF/COPD/diabetes	Urethral stenosis	31	5	Moderate	1	Yellow	• SSIH: Blue • SSIR: Green • SSIP: — • SSID: Green • SSIT: — • SSIM: Green
Male/64 years	CHF/COPD/diabetes	Arterial rupture	42	1	Moderate	5[p]	Orange	• SSIH: Green • SSIR: Orange • SSIP: Orange • SSID: Yellow • SSIT: Green • SSIM: Green

[a]Serious adverse events were coded using the MedDRA dictionary version 19.1.

[b]ORI: overall risk indicator.

[c]Last color of ORI is a weighted combination of the SSI status with specific calculation rules in case of missing data, depending on the time interval since their last available transmission.

[d]Green ORI, no action required; yellow ORI, low-priority action; orange ORI, medium-priority action; and red ORI, high-priority action.

[e]SSI: stability system indicator.

[f]Green SSIs indicate stability, whereas blue, yellow, orange, and red SSIs indicate decreasing stability, from slightly to markedly unstable. Blue SSI triggers increased frequency of data measurements.

[g]CHF: chronic heart failure.

[h]SSIH: hemodynamic stability system indicator.

[i]SSIR: respiratory stability system indicator.

[j]SSIP: pain/sleep quality stability system indicator.

[k]SSID: diet/diuresis stability system indicator.

[l]SSIT: tissue integrity stability system indicator.

[m]SSIM: medication stability system indicator.

[n]Patient was in the NOMHAD suspension period at his own request.

[o]Not available.

[p]Patient was in the NOMHAD suspension period due to programmed hospital admission.

Discussion

Principal Findings

This exploratory study demonstrated that the NOMHAD eHealth system was accurate in generating ORIs that reflect health status changes in patients with multimorbidity. The ORIs generated by the system agreed well with the ORIs that were expected when they were recalculated using the raw data teletransmitted. Sensitivity and specificity were both high for the ORIs and for the individual clinical stability system indicators. These results indicate that the system provided an efficient and accurate means of transmitting health status data and that the ORIs generated by the algorithm accurately reflected the health stability of the patients.

All study patients generated telenotifications, most of which were of low or moderate priority. Less than one-third of the patients required a high-priority telecoaching action. However, because ORIs were systematically recalculated every time new data were teletransmitted, a high number of ORIs (N=6263) were generated. This total was almost 3 times higher than what was anticipated for 23 patients over the study duration (100 days), assuming 1 telenotification per patient per day (~2300 ORIs). This disparity reflected the conservative nature in which the telenotifications were programmed to assure the safety of these patients with multimorbidity, which was a main priority during this study. In addition, each time a patient variable was out of the expected normal range, the self-measurement had to be repeated. Another reason for the high number is that some patients became high consumers of the system and performed their own repeated measurements every day. This could also explain why 71.53% (4480/6263) of the ORIs were of low (yellow) priority. A high number of ORIs were generated, leading to a substantial nurse workload, as a median of 21 outgoing telenotification management–related contacts were made to each patient over the 3-month follow-up. In managing the telenotifications, the call center nurses were able to effectively filter those requiring high-priority telecoaching actions. It is anticipated that future technical developments will reduce the number of low-priority ORIs.

The study showed that the NOMHAD eHealth system was generally well accepted by patients, nurses, and physicians, all of whom indicated high levels of satisfaction with most aspects of the program and a willingness to use the system long term. Responses to the self-questionnaires indicated that the system achieved a key aim of telehealth programs: to better manage and coordinate timely care for patients with multiple chronic diseases. Importantly, most users felt that it helped patients to better manage their diseases and symptoms and better understand their illnesses. Nurses felt that it helped them to communicate with patients more effectively, whereas most physicians felt that it helped them to better manage their patients. Most of the physicians also felt that the patients seen in consultation were only those who truly needed physician intervention. For patients, regular contact with the call center nurses was considered an important aspect of the program, and many of them considered this as the best part. Chronic diseases often limit social contact, especially for patients who live alone. It appears that contact with the nurses provided a key social interaction for many patients.

Patient satisfaction and participation were lower for the online video classroom sessions. Not all of the 8 different 40-minute sessions were relevant for every patient and attendance was relatively low. In addition, the mountainous terrain of the study setting caused some technical and connectivity issues, which affected the transmission quality of the videoconferences and may have frustrated some patients. By contrast, patients consulted the disease self-management support modules provided on their tablet approximately 11 times, indicating that they were interested in learning about their chronic diseases. These support modules were likely considered more convenient because patients could consult these at any time, unlike the online video classroom sessions, which were scheduled weekly at fixed times.

A recent study also investigated the views and experiences of patients and health care personnel using a telehealth and online self-management program for patients with multiple chronic diseases (diabetes, cardiovascular disease, and hypertension) in the Renewing Health Project [33]. Although there were differences between the 2 programs, both focused on patient-reported health status variables and on improving self-management skills. An important finding was that the expectations of patients, who expected feedback and support, differed from those of the health care staff, who expected the program to make patients more independent and able to perform self-management tasks without staff support. By contrast, the regular every 2-week patient–nurse contact, which provided consistent feedback and support, was a key aspect of the NOMHAD program and highly regarded by the patients. Together, these findings suggest that telehealth programs should provide a means of maintaining a level of personal contact that preserves patient engagement. Insufficient contact may result in patients feeling isolated or abandoned.

XSL•FO

RenderX

Strengths

This study has several strengths. Whereas most previous telehealth programs have focused on a single disease, the NOMHAD eHealth system was designed for patients with multiple chronic diseases. With multimorbidity on the rise, the study patients likely represent a more realistic "real-world" situation than those studied in many other telehealth programs. In addition, patients were living in a rural mountainous area, whereas the call center nurses were nearly 750 km away and the physicians were in private practice or a hospital nearby. Despite some connectivity issues, which are to be expected in such settings, this study demonstrated the system's feasibility in remote conditions where many patients may face medical isolation. The telenotification accuracy of the system was verified, both for overall assessments of health status and for the individual clinical stability systems indicators. In addition, the NOMHAD eHealth system collects a comprehensive array of patient data that include subjective assessments of well-being, as well as objective self-measurements of patient health variables.

Limitations

Regarding limitations, the self-questionnaires were created specifically for this study and were not independently validated. Besides, very few nurses and physicians participated in the study. Therefore, the questionnaire results for these health care providers must be viewed with caution. Although the feasibility results were positive, the follow-up period was relatively short for a program designed to be used long term. These results need to be confirmed in a longer study. Other minor problems were that the standard glycemia thresholds were too narrow, which resulted in irrelevant notifications, and that the blood pressure cuffs were too small for some patients with obesity.

The NOMHAD eHealth system was derived from earlier Spanish versions of the NOMHAD Chronic software platform (version 1.5.2) and the NOMHAD Mobile app (version 1.2.0). The Spanish version was used in a 3-arm randomized clinical trial comparing a structured telephone intervention, with and without the NOMHAD eHealth system, with usual care in 472 elderly patients with multimorbidity [31]. In Spain, the telemonitoring was performed by a hospital nurse team, whereas in this study in France, it was based on call center nurses. After a 12-month follow-up, the main outcome of health-related quality of life was significantly higher with the telehealth intervention than with usual care ($P<.001$). No differences in mortality or health care utilization were found. Nevertheless, this earlier study also demonstrated that the NOMHAD eHealth system was feasible for long-term use [31].

The medical relevance of the medium- to high-priority telenotifications (orange and red ORIs) also needs to be evaluated because of the significant workload they generated for nurses. As the aim of the study was to support high-risk patients, those included in this study were highly comorbid, with a mean of 6.5 concomitant diseases, and had severe targeted diseases, documented by a low left ventricular ejection fraction, a low forced expiratory volume in 1 second, and high proportions of insulin-dependent patients or those requiring home oxygen or ventilator support. Moreover, 7 patients underwent an unscheduled acute hospital admission during their 3-month follow-up. This preliminary study was not designed to assess medical relevance of the generated ORIs or clinical benefits, which would require a much larger and longer randomized controlled trial. However, several recent studies of similar telehealth and telemonitoring programs in patients with multimorbidity have reported positive results, suggesting that such programs can reduce disease exacerbations and health care costs [34,35]. A program for patients with chronic conditions in Australia found that, compared with the previous year, the telemonitoring program reduced medical expenses by 46.3%, pharmaceutical expenses by 25.5%, unscheduled hospital admissions by 53.2%, hospital length of stay by 67.9%, and mortality by over 40% [34]. In Spain, the ValCrònic program for patients with multiple chronic conditions found significant reductions in body weight; significantly smaller proportions of patients with high systolic or diastolic blood pressure, or elevated glycosylated hemoglobin concentrations in patients with diabetes; and significant reductions in emergency service use and hospitalizations [35].

Conclusions

In conclusion, this study met its principal objectives by providing evidence that the NOMHAD eHealth system is accurate, useful, acceptable, informative, and feasible to implement and use. These favorable outcomes depended on the commitment and coordination of all users and study staff, all of which were facilitated by the system network. Although the NOMHAD eHealth system is based on efficient teletransmission of electronic data and accurate algorithmic evaluation, the system also facilitates human interaction on a regular basis, which appeared to be one of its most important features. The system was also considered useful to support disease education and promote patient awareness toward a better understanding of health data. This study confirms that the system is feasible for larger, longer studies to investigate its impacts on clinical variables, patient-reported outcomes, and medical resources used. However, in their comprehensive meta-review, Elbert et al [36] concluded that, given the consistent effectivity of telehealth for most patients, greater attention should be given to intervention evaluation rather than to obtaining additional evidence for their efficacy.

XSL·FO

RenderX

Authors' Contributions

All authors contributed to the study design and writing of the protocol. MC, AF, and SH contributed to data collection. All authors contributed to data analysis and interpretation, contributed to the writing or reviewing of the manuscript, and approved the final version.

Conflicts of Interest

JT is an employee of VitalAire France, a home health care provider, an Air Liquide health care subsidiary. CB is an employee of Air Liquide Santé International. MC, AF, and SH received honoraria from Air Liquide Santé International for the conduct of the trial.

References

1. GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016 Oct 8;388(10053):1545-1602 [FREE Full text] [doi: 10.1016/S0140-6736(16)31678-6] [Medline: 27733282]
2. The top 10 causes of death. World Health Organization. 2016. URL: http://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [accessed 2018-12-17]
3. Agence nationale d'appui à la performance des établissements de santé et médico sociaux. La télémédecine en action: 25 projets passés à la loupe. Un éclairage pour le déploiement national. Tome 1: les grands enseignements. Paris, France: ANAP; 2012. URL: http://www.anap.fr/ressources/publications/detail/actualites/la-telemedecine-en-action-25-projets-passes-a-la-loupe/ [accessed 2021-07-09]
4. Lutte contre les déserts médicaux : ce qu'il faut savoir (Battle against medical isolation: what is necessary to know). French Government. URL: https://www.gouvernement.fr/lutte-contre-les-deserts-medicaux-ce-qu-il-faut-savoir [accessed 2021-07-09]
5. Bloom D, Cafiero E, Jané-Llopis E, Abrahams-Gessel S, Bloom L, Fathima S. The Global Economic Burden of Noncommunicable Diseases. Program on the Global Demography of Aging. 2012. URL: https://EconPapers.repec.org/RePEc:gdm:wpaper:8712 [accessed 2021-07-09]
6. World Health Organization. Scaling up action against noncommunicable diseases: how much will it cost? WHO. Geneva: World Health Organization; 2011. URL: http://apps.who.int/iris/bitstream/handle/10665/44706/9789241502313_eng.pdf?sequence=1 [accessed 2018-12-17]
7. Conway A, Inglis SC, Chang AM, Horton-Breshears M, Cleland JGF, Clark RA. Not all systematic reviews are systematic: a meta-review of the quality of systematic reviews for non-invasive remote monitoring in heart failure. J Telemed Telecare 2013 Sep;19(6):326-337. [doi: 10.1177/1357633X13503427] [Medline: 24163297]
8. Wilkinson TMA, Donaldson GC, Hurst JR, Seemungal TAR, Wedzicha JA. Early therapy improves outcomes of exacerbations of chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2004 Jun 15;169(12):1298-1303. [doi: 10.1164/rccm.200310-1443OC] [Medline: 14990395]
9. Bashi N, Karunanithi M, Fatehi F, Ding H, Walters D. Remote Monitoring of Patients With Heart Failure: An Overview of Systematic Reviews. J Med Internet Res 2017 Jan 20;19(1):e18 [FREE Full text] [doi: 10.2196/jmir.6571] [Medline: 28108430]
10. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. N Engl J Med 2010 Dec 9;363(24):2301-2309 [FREE Full text] [doi: 10.1056/NEJMoa1010029] [Medline: 21080835]
11. Nakamura N, Koga T, Iseki H. A meta-analysis of remote patient monitoring for chronic heart failure patients. J Telemed Telecare 2014 Jan;20(1):11-17. [doi: 10.1177/1357633X13517352] [Medline: 24352899]
12. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telemonitoring for congestive heart failure: a systematic review and meta-analysis. J Telemed Telecare 2010;16(2):68-76. [doi: 10.1258/jtt.2009.090406] [Medline: 20008054]
13. Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Kirwan B, et al. Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial. Lancet 2018 Dec 22;392(10152):1047-1057 [FREE Full text] [doi: 10.1016/S0140-6736(18)31880-4] [Medline: 30153985]

14. Kotb A, Cameron C, Hsieh S, Wells G. Comparative effectiveness of different forms of telemedicine for individuals with heart failure (HF): a systematic review and network meta-analysis. PLoS One 2015;10(2):e0118681 [FREE Full text] [doi: 10.1371/journal.pone.0118681] [Medline: 25714962]

15. Feltner C, Jones CD, Cené CW, Zheng Z, Sueta CA, Coker-Schwimmer EJL, et al. Transitional care interventions to prevent readmissions for persons with heart failure: a systematic review and meta-analysis. Ann Intern Med 2014 Jun 3;160(11):774-784. [doi: 10.7326/M14-0083] [Medline: 24862840]

16. Kessler R, Casan-Clara P, Koehler D, Tognella S, Viejo JL, Dal Negro RW, et al. COMET: a multicomponent home-based disease-management programme routine care in severe COPD. Eur Respir J 2018 Jan;51(1):1701612 [FREE Full text] [doi: 10.1183/13993003.01612-2017] [Medline: 29326333]

17. Chatwin M, Hawkins G, Panicchia L, Woods A, Hanak A, Lucas R, et al. Randomised crossover trial of telemonitoring in chronic respiratory patients (TeleCRAFT trial). Thorax 2016 Apr;71(4):305-311 [FREE Full text] [doi: 10.1136/thoraxjnl-2015-207045] [Medline: 26962013]

18. Walker PP, Pompilio PP, Zanaboni P, Bergmo TS, Prikk K, Malinovschi A, et al. Telemonitoring in Chronic Obstructive Pulmonary Disease (CHROMED). A Randomized Clinical Trial. Am J Respir Crit Care Med 2018 Sep 01;198(5):620-628. [doi: 10.1164/rccm.201712-2404OC] [Medline: 29557669]

19. Pedone C, Lelli D. Systematic review of telemonitoring in COPD: an update. Pneumonol Alergol Pol 2015;83(6):476-484 [FREE Full text] [doi: 10.5603/PiAP.2015.0077] [Medline: 26559801]

20. Cruz J, Brooks D, Marques A. Home telemonitoring effectiveness in COPD: a systematic review. Int J Clin Pract 2014 Mar;68(3):369-378. [doi: 10.1111/ijcp.12345] [Medline: 24472009]

21. Mirón Rubio M, Ceballos Fernández R, Parras Pastor I, Palomo Iloro A, Fernández Félix BM, Medina Miralles J, et al. Telemonitoring and home hospitalization in patients with chronic obstructive pulmonary disease: study TELEPOC. Expert Rev Respir Med 2018 Apr;12(4):335-343. [doi: 10.1080/17476348.2018.1442214] [Medline: 29460648]

22. Lundell S, Holmner, Rehn B, Nyberg A, Wadell K. Telehealthcare in COPD: a systematic review and meta-analysis on physical outcomes and dyspnea. Respir Med 2015 Jan;109(1):11-26. [doi: 10.1016/j.rmed.2014.10.008] [Medline: 25464906]

23. Goldstein RS, O'Hoski S. Telemedicine in COPD: time to pause. Chest 2014 May;145(5):945-949. [doi: 10.1378/chest.13-1656] [Medline: 24798834]

24. Marcolino MS, Maia JX, Alkmim MBM, Boersma E, Ribeiro AL. Telemedicine application in the care of diabetes patients: systematic review and meta-analysis. PLoS One 2013;8(11):e79246 [FREE Full text] [doi: 10.1371/journal.pone.0079246] [Medline: 24250826]

25. McDonnell ME. Telemedicine in Complex Diabetes Management. Curr Diab Rep 2018 May 24;18(7):42. [doi: 10.1007/s11892-018-1015-3] [Medline: 29797292]

26. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K. Home telehealth for diabetes management: a systematic review and meta-analysis. Diabetes Obes Metab 2009 Oct;11(10):913-930. [doi: 10.1111/j.1463-1326.2009.01057.x] [Medline: 19531058]

27. Walker CL, Kopp M, Binford RM, Bowers CJ. Home Telehealth Interventions for Older Adults With Diabetes. Home Healthc Now 2017 Apr;35(4):202-210. [doi: 10.1097/NHH.0000000000000522] [Medline: 28353510]

28. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 2012 Jul 7;380(9836):37-43 [FREE Full text] [doi: 10.1016/S0140-6736(12)60240-2] [Medline: 22579043]

29. Lefèvre T, d'Ivernois J, De Andrade V, Crozet C, Lombrail P, Gagnayre R. What do we mean by multimorbidity? An analysis of the literature on multimorbidity measures, associated factors, and impact on health services organization. Rev Epidemiol Sante Publique 2014 Oct;62(5):305-314 [FREE Full text] [doi: 10.1016/j.respe.2014.09.002] [Medline: 25444838]

30. Onder G, Palmer K, Navickas R, Jurevičienė E, Mammarella F, Strandzheva M, Joint Action on Chronic DiseasesPromoting Healthy Ageing across the Life Cycle (JA-CHRODIS). Time to face the challenge of multimorbidity. A European perspective from the joint action on chronic diseases and promoting healthy ageing across the life cycle (JA-CHRODIS). Eur J Intern Med 2015 Apr;26(3):157-159. [doi: 10.1016/j.ejim.2015.02.020] [Medline: 25797840]

31. Valdivieso B, García-Sempere A, Sanfélix-Gimeno G, Faubel R, Librero J, Soriano E, GeChronic Group. The effect of telehealth, telephone support or usual care on quality of life, mortality and healthcare utilization in elderly high-risk patients with multiple chronic conditions. A prospective study. Med Clin (Barc) 2018 Oct 23;151(8):308-314. [doi: 10.1016/j.medcli.2018.03.013] [Medline: 29705155]

32. Haute Autorité de Santé. URL: https://www.has-sante.fr/portail/jcms/fc_1249605/fr/outils-guides-methodes [accessed 2021-07-09]

33. Lindberg I, Lindberg B, Söderberg S. Patients' and Healthcare Personnel's Experiences of Health Coaching with Online Self-Management in the Renewing Health Project. Int J Telemed Appl 2017;2017:9306192 [FREE Full text] [doi: 10.1155/2017/9306192] [Medline: 29463979]

34. Celler B, Varnfield M, Nepal S, Sparks R, Li J, Jayasena R. Impact of At-Home Telemonitoring on Health Services Expenditure and Hospital Admissions in Patients With Chronic Conditions: Before and After Control Intervention Analysis. JMIR Med Inform 2017 Sep 08;5(3):e29 [FREE Full text] [doi: 10.2196/medinform.7308] [Medline: 28887294]

35. Orozco-Beltran D, Sánchez-Molla M, Sanchez JJ, Mira JJ, ValCrònic Research Group. Telemedicine in Primary Care for Patients With Chronic Conditions: The ValCrònic Quasi-Experimental Study. J Med Internet Res 2017 Dec 15;19(12):e400 [FREE Full text] [doi: 10.2196/jmir.7677] [Medline: 29246881]
36. Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, Hakkaart-van Roijen L, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr 16;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]

Abbreviations

AE: adverse event
CHF: chronic heart failure
COPD: chronic obstructive pulmonary disease
ORI: overall risk indicator
SSI: stability system indicator

XSL•FO
RenderX

Original Paper

CoGNIT Automated Tablet Computer Cognitive Testing in Patients With Mild Cognitive Impairment: Feasibility Study

Anders Behrens[1], MD, PhD; Johan Sanmartin Berglund[1], MD, PhD; Peter Anderberg[1], MSc, PhD

Department of Health, Blekinge Institute of Technology, Karlskrona, Sweden

Corresponding Author:
Peter Anderberg, MSc, PhD
Department of Health
Blekinge Institute of Technology
Valhallavägen 1
Karlskrona, 37141
Sweden
Phone: 46 734223736
Email: pan@bth.se

Abstract

Background: Early diagnosis of cognitive disorders is becoming increasingly important. Limited resources for specialist assessment and an increasing demographical challenge warrants the need for efficient methods of evaluation. In response, CoGNIT, a tablet app for automatic, standardized, and efficient assessment of cognitive function, was developed. Included tests span the cognitive domains regarded as important for assessment in a general memory clinic (memory, language, psychomotor speed, executive function, attention, visuospatial ability, manual dexterity, and symptoms of depression).

Objective: The aim of this study was to assess the feasibility of automatic cognitive testing with CoGNIT in older patients with symptoms of mild cognitive impairment (MCI).

Methods: Patients older than 55 years with symptoms of MCI (n=36) were recruited at the research clinic at the Blekinge Institute of Technology (BTH), Karlskrona, Sweden. A research nurse administered the Mini-Mental State Exam (MMSE) and the CoGNIT app on a tablet computer. Technical and testing issues were documented.

Results: The test battery was completed by all 36 patients. One test, the four-finger–tapping test, was performed incorrectly by 42% of the patients. Issues regarding clarity of instructions were found in 2 tests (block design test and the one finger-tapping test). Minor software bugs were identified.

Conclusions: The overall feasibility of automatic cognitive testing with the CoGNIT app in patients with symptoms of MCI was good. The study highlighted tests that did not function optimally. The four-finger–tapping test will be discarded, and minor improvements to the software will be added before further studies and deployment in the clinic.

(JMIR Form Res 2022;6(3):e23589)   doi:10.2196/23589

KEYWORDS

internet; cognitive testing; software; testing; impairment; cognition; feasibility; diagnosis; app; assessment; cognitive impairment

Introduction

The global burden of dementia is estimated to increase from 43-47 million in 2016 to over 100 million patients by 2050 [1]. Early diagnosis of cognitive diseases is of increasing importance to assure optimal and early care, prognostics, and treatment. This is highlighted by the development of criteria for mild cognitive impairment (MCI), in which diagnostic criteria exist today for predementia stages of Alzheimer disease, dementia in Parkinson disease, and vascular cognitive impairment. Early detection might be even more important in the future when more potent treatments might exist that are likely to be more effective at an early stage of disease.

In the latest revision of the American Academy of Neurology practice guidelines for MCI (September 2019), yearly cognitive screening in healthy adults above 65 years old is encouraged [2]. Longitudinal assessments of patients with MCI are even more important, as approximately 50% might have a cognitive disease with progressive cognitive decline, and many patients revert to normal cognitive functioning [3,4]. This predicts an increasing demand for cognitive testing. Commonly used screening instruments such as the Mini-Mental State Exam

XSL•FO
RenderX

(MMSE) are rapid but crude, and measurement error is larger than the typical yearly decline in Alzheimer disease [5,6]. Also, the utility for differential diagnostics is limited. Neuropsychological assessment with sensitive measures of several cognitive domains may aid diagnostics when basal investigation including screening instruments are inconclusive.

Today, less than one-third of patients in Sweden referred from primary to specialist care centers undergo a neuropsychological investigation. Given the demographical challenge, this portion is likely to decrease further. In response, CoGNIT, a computerized neuropsychological test battery, was developed, initially for assessing hydrocephalus patients. Tests are automatically delivered and scored on a tablet computer, assuring portability, standardization, and efficiency. Made for repeated assessments, included tests are chosen to be able to capture performance on a spectrum, rather than dichotomous variables, thus being able to pick up nuances in cognitive change. The test battery includes tests to cover several cognitive domains (memory, visuospatial function, attention, psychomotor speed, executive function, and manual dexterity). A summary score is calculated for all tests, making it possible to track changes in specific domains as well as global functioning over time.

A previous version of CoGNIT that ran on a touch-screen computer has been validated, norms and reliability data have been collected, and the test has been used in clinical practice in the management of hydrocephalus patients for over 5 years [7,8]. A novel version of CoGNIT running on a tablet has been expanded with tests of language, visuospatial ability, and upper extremity motor speed, to cover the cognitive domains of a typical evaluation in a memory clinic. The intended use cases are (1) as part of the investigation at a memory clinic for early detection of cognitive impairment; (2) in differential diagnostics, since diseases have different patterns of cognitive impairment; and (3) in follow-up testing after initiation of treatment (eg, acetylcholinesterase inhibitor) to assess the treatment effect.

The aim of this study was to assess the feasibility of automatic testing with CoGNIT in older persons with MCI. Data will be used to adapt the software for better performance before collecting normative data in a healthy population.

Methods

Ethical Approval

Ethical approval for this study was granted by the regional ethical review board in Lund, Sweden (dnr 2016/470).

Recruitment

Patients with symptoms of MCI were recruited from the ongoing SMART4MD project at the research clinic at the Blekinge Institute of Technology (BTH), Karlskrona, Sweden [9]. This project evaluates a digital platform that supports patients with reminders, information, and memory support in everyday life. Patients in this project have several visits to the research clinic during their follow-up and were recruited for cognitive testing with CoGNIT during their second visit. The examiner (a research nurse) explained the experiment and collected signed consent from each subject. Patients were included in February 2020 and March 2020. Inclusion criteria were (data collected at first visit) MMSE of 20 to 28 points; subjective memory problems for >6 months; older than 55 years; participants in charge over their own medication; and no specific conditions reducing their ability to use an app, such as visual, hearing, or motor impairments.

Exclusion criteria were a terminal illness with less than 3 years of expected survival, Geriatric Depression Scale (GDS) score above 11, or cognitive impairment due to a known condition such as abuse or psychiatric illness.

Procedures

The MMSE was administered to participants by the research nurse (the examiner). In direct succession, the CoGNIT computerized neuropsychological test battery was administered. While testing, the examiner was sitting by, answering questions, and taking notes regarding technical or testing issues. Patients were also asked to answer the TechPH questionnaire, a short instrument assessing older people's attitude towards technology [10].

CoGNIT Test Battery

Test Presentation and Instructions

Tests were presented on a 10.5" tablet computer (Samsung Galaxy Tab S6, 6 GB RAM, 128 GB internal storage; Samsung Electronics, South Korea) running the Android 11.0 operating system. The tablet was connected via Wi-Fi to the internet. CoGNIT was accessed on the Chrome web browser via an online server. Results from testing were stored in a database on the server for access via a web interface. Test instructions are presented with animations and sounds via a speaker. A trial round with automatic feedback precedes all tests. Two tests require verbal input from the patient that is recorded by the tablet microphone (10-word list test and category fluency test). These tests are manually corrected by the examiner after the testing session (simple procedure by "checking boxes" in the software). All other tests are automatically scored by the software. When scoring is completed, a pdf test report is automatically produced. Included tests are described in the following sections.

Memory

In the 10-word list test, 10 nouns are consecutively presented via text and sound. The patient is asked to say the remembered words aloud after presentation. The test is repeated for 3 trials. The sound is recorded for later manual scoring of the test. After approximately 15 minutes and 2 intervening distractor tasks, a free recall test is performed, without prior presentation of the words. In direct succession, a recognition test is performed. The patient is asked to press yes/no buttons indicating if the word was included in the learning test. A total of 20 words is presented: 10 correct and 10 distractor words.

Executive Function and Psychomotor Speed

The Stroop test consists of 2 parts, the Stroop congruent test and Stroop incongruent test. In the Stroop congruent test, the patient is asked to press 1 of 2 colored buttons on the screen indicated by text. In the Stroop incongruent test, the text is colored, and the patient is asked to press the button indicated

by the color of the text and not what is written. Each trial includes 50 color-words. Response time and errors are collected.

Similarly, the Trail Making Test also includes 2 tests. In Trail Making Test A, the numbers 1 to 25 are presented on the screen, and the patient is asked to press each number in numerical order. In Trail Making Test B, both numbers (1-13) and letters (A-L) are presented on the screen. The patient is asked to press letters and number in order by alternating between successive numbers and letters of the alphabet (ie, 1-A-2-B-3-C...). Time to completion and errors are collected.

Attention

In the 2-choice reaction test, the patient is asked to press 1 of 2 buttons indicated by an arrow. The arrow appears after a random interval of 5 seconds to 15 seconds. Reaction time is measured for 20 trials.

Language

In the category fluency test, the patient is asked to say aloud as many words from a category (eg, animals) as possible for 1 minute. Sound is recorded and corrected by the examiner after all tests are completed.

Motor Speed

In the four-finger–tapping test, the patient is required to tap on a small keyboard with the digits 2 through 5. The correct order of taps is (digits) 2-3-4-5-4-3-2-3... The number of correct taps over 5 trials is scored. In the one-finger–tapping test, the patient taps 1 finger repeatedly between 2 circles on the screen. The number of taps during a 10-second period is recorded for 3 trials for both the left and right hands.

Visuospatial Ability

The block design test, a figure composed of 4 colored blocks is presented on the screen. The patient is asked to rearrange blocks presented on the top of the screen to match the pattern. Time to completion is scored for 3 different patterns.

Depression

Symptoms of depression were screened using the GDS, in which 15 questions are presented on the screen and the patient is required to answer by pressing buttons marked yes or no.

Statistical Analysis

The completion rate without testing issues in patients was previously estimated at 80% in hydrocephalus patients. With a sample size of 36, an 80% issue-free completion rate can be estimated with a 95% CI of +/-13%, which was regarded as sufficient. This also assured for detecting issues for improvement. Influences on test completion by age and score on the TechPH and MMSE scales were assessed with Mann-Whitney U tests. Influence by gender and education was assessed with chi-squared tests. Significance level was set at $P<.05$. All statistics were analyzed in SPSS (Version 25; IBM Corp, Armonk, NY).

Results

Sample Characteristics

Of 42 patients, 36 (86%) agreed to participate; 4 patients declined participation due to a lack of interest or time, and 2 patients declined because of medical reasons (hemiparesis and severe visual impairment). Demographics for the study population are presented in Table 1. The age was rather high (mean 75.6 years). There was a slight overrepresentation of men (22/36, 61%).

Table 1. Demographic data for the study population.

Characteristics	Results
Age, (years), mean (SD)	75.6 (5.0)
Sex, n (%)	
Male	22 (61)
Female	14 (39)
Education, n (%)	
Low	9 (25)
Medium	14 (39)
High	13 (36)
MMSE[a] (points), mean (SD)	28.2 (2.1)
TechPH[b] (points), mean (SD)	3.0 (0.8)

[a]MMSE: Mini-Mental State Exam.

[b]TechPH: novel scale assessing older people's attitude towards technology. The score range is 1-5. Higher scores indicate higher level of technophilia.

Testing Issues

Testing issues are summarized in Table 2. Excluding failure due to technical or physical issues impairing the ability to complete a test, failure to complete a test due to

misunderstanding of instructions was observed for 11 patients (10 four-finger–tapping test, 2 one-finger–tapping, and 1 block design test). Age, gender, education level, MMSE, and TechPH scores were not associated with failure. The patient in the cohort

with the lowest MMSE (21 points) misunderstood instructions in both the four- and one-finger–tapping tests.

Table 2. Testing issues identified by the examiner.

Test	Issue
Four-finger tapping	Of the 36 patients, 15 (42%) did not perform the test as intended because of physical difficulties (rheumatoid arthritis, fractures, or arthrosis) or misunderstanding of instructions; 3 patients did not comprehend how to use the keypad and instead tried to perform the test on the screen of the tablet.
Stroop test	A yellow button was perceived as light green by 2 patients.
Block design test	One patient did not understand when the test started.
Geriatric Depression Scale	Two patients gave a wrong answer to 1 question. Once answered it was not possible to correct.
One-finger tapping	One patient performed all tests with the left hand. One patient needed extra instructions from the examiner regarding that the buttons were supposed to be tapped and not to drag the finger between buttons.

Technical Issues

We observed 2 technical issues. In 3 test sessions of the 10-word list test, 1 instruction was repeated several times. For 1 testing session, 2 pdf reports were created.

Discussion

Principal Findings

We developed CoGNIT, a tablet app for assessing cognitive function in several cognitive domains in an automatic, standardized, and efficient manner. A previous version was developed for assessment of hydrocephalus patients, which has now been redesigned for tablet computers and expanded with new tests for more general assessments, aimed at aiding diagnostics and tracking of cognitive function in patients at memory clinics. This study aimed to assess the feasibility of testing patients with symptoms of MCI.

Overall, the feasibility for testing with patients with symptoms of MCI was good, although some issues were identified. Software bugs were identified that are straightforward to correct. Some tests did not perform optimally. On a group level, there was no indication that this was due to a lower level of a technophilia personality trait or cognitive ability as measured by the TechPH and MMSE scales. However, cognitive ability might interfere with testing in the lower range of the MMSE, as the patient with the lowest MMSE in the cohort failed to complete 2 of the included tests due to miscomprehension of instructions.

Most notably, 42% of the patients did not perform the four-finger–tapping test as intended, both due to physical limitations (eg, rheumatoid arthritis) and difficulty understanding instructions. Though a measure of manual dexterity, scores from this test correlate with tests of executive function [11]. Tests of executive function have a component of "getting it," and this test might just be too cognitively demanding for the MCI population. Also, introducing a second means of input from the attached keyboard might be confusing. The test was originally included in the CoGNIT battery because of evidence for assessing patients with hydrocephalus [12]. A more prevalent

disease in memory clinics is any of the parkinsonian syndromes, where the one-finger–tapping test has evidence [3,13]. After evaluation, the four-finger–tapping test was discarded in favor of this test in further deployment of CoGNIT.

Other testing issues are more straightforward to improve. The block design and the one-finger–tapping tests will be improved with updated instructions for the tests. In the Stroop test, a simple color adjustment will be done to improve discrimination between yellow and green. In the GDS, a button to access the previous question will be added.

Limitations

Though inclusion criteria at the first visit included an MMSE score in the range of 20 to 28, many patients scored higher at the second visit, when the CoGNIT was administered. There was a skewed distribution, with many patients scoring in the higher range. The feasibility for testing with more cognitively impaired patients is thus less tested. However, sensitive neuropsychological testing is needed less in the cognitive range where screening instruments show marked impairment.

Comparison With Prior Work

There are several computerized test batteries for neuropsychological testing of older adults [14]. Most batteries are administered by a trained testing technician who explains instructions for each test. This hampers standardization and scalability. A major criticism of computerized tests has been a lack of reports of reliability, normative data, validity, and, finally, poorly designed computer-person interfaces. All these issues were addressed when designing CoGNIT. Also, testing free recall memory is, to our knowledge, unique to CoGNIT. Free recall is the most sensitive test for disorders of episodic memory. We believe this will give CoGNIT an edge in early detection of cognitive disorders.

Conclusions

The feasibility for automatic neuropsychological testing with the CoGNIT app on a tablet device in patients with symptoms of MCI was good. After minor modifications, the app is ready for further studies. The next step is collecting normative data from a healthy population.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Korsnes MS, Winkler AS. Global, regional, and national burden of dementia, 1990-2016: Predictions need local calibration. Neurology 2020 Apr 21;94(16):718-719. [doi: 10.1212/WNL.0000000000009301] [Medline: 32245845]

2. Foster NL, Bondi MW, Das R, Foss M, Hershey LA, Koh S, et al. Quality improvement in neurology: Mild cognitive impairment quality measurement set. Neurology 2019 Oct 15;93(16):705-713. [doi: 10.1212/WNL.0000000000008259] [Medline: 31534026]

3. Palmqvist S, Hertze J, Minthon L, Wattmo C, Zetterberg H, Blennow K, et al. Comparison of brief cognitive tests and CSF biomarkers in predicting Alzheimer's disease in mild cognitive impairment: six-year follow-up study. PLoS One 2012;7(6):e38639 [FREE Full text] [doi: 10.1371/journal.pone.0038639] [Medline: 22761691]

4. Overton M, Pihlsgård M, Elmståhl S. Diagnostic stability of mild cognitive impairment, and predictors of reversion to normal cognitive functioning. Dement Geriatr Cogn Disord 2019;48(5-6):317-329 [FREE Full text] [doi: 10.1159/000506255] [Medline: 32224608]

5. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]

6. Hensel A, Angermeyer MC, Riedel-Heller SG. Measuring cognitive change in older adults. Do reliable change indices of the SIDAM predict dementia? J Neurol 2007 Oct;254(10):1359-1365. [doi: 10.1007/s00415-007-0549-0] [Medline: 17932704]

7. Behrens A, Eklund A, Elgh E, Smith C, Williams MA, Malm J. A computerized neuropsychological test battery designed for idiopathic normal pressure hydrocephalus. Fluids Barriers CNS 2014;11:22 [FREE Full text] [doi: 10.1186/2045-8118-11-22] [Medline: 25279138]

8. Behrens A, Elgh E, Leijon G, Kristensen B, Eklund A, Malm J. The Computerized General Neuropsychological INPH Test revealed improvement in idiopathic normal pressure hydrocephalus after shunt surgery. J Neurosurg 2019 Feb 08;132(3):733-740. [doi: 10.3171/2018.10.JNS18701] [Medline: 30738407]

9. Anderberg P, Barnestein-Fonseca P, Guzman-Parra J, Garolera M, Quintana M, Mayoral-Cleries F, et al. The effects of the digital platform Support Monitoring and Reminder Technology for Mild Dementia (SMART4MD) for people with mild cognitive impairment and their informal carers: protocol for a pilot randomized controlled trial. JMIR Res Protoc 2019 Jun 21;8(6):e13711 [FREE Full text] [doi: 10.2196/13711] [Medline: 31228177]

10. Anderberg P, Eivazzadeh S, Berglund JS. A novel instrument for measuring older people's attitudes toward technology (TechPH): development and validation. J Med Internet Res 2019 May 23;21(5):e13951 [FREE Full text] [doi: 10.2196/13951] [Medline: 31124467]

11. Behrens A. Measurements in idiopathic normal pressure hydrocephalus: Computerized neuropsychological test battery and intracranial pulse waves. Umeå University. 2014. URL: http://umu.diva-portal.org/smash/get/diva2:762579/FULLTEXT01.pdf [accessed 2022-02-27]

12. Lenfeldt N, Larsson A, Nyberg L, Andersson M, Birgander R, Eklund A, et al. Idiopathic normal pressure hydrocephalus: increased supplementary motor activity accounts for improvement after CSF drainage. Brain 2008 Nov;131(Pt 11):2904-2912. [doi: 10.1093/brain/awn232] [Medline: 18931387]

13. Lee CY, Kang SJ, Hong S, Ma H, Lee U, Kim YJ. A validation study of a smartphone-based finger tapping application for quantitative assessment of bradykinesia in Parkinson's disease. PLoS One 2016;11(7):e0158852 [FREE Full text] [doi: 10.1371/journal.pone.0158852] [Medline: 27467066]

14. Zygouris S, Tsolaki M. Computerized cognitive testing for older adults: a review. Am J Alzheimers Dis Other Demen 2015 Feb;30(1):13-28 [FREE Full text] [doi: 10.1177/1533317514522852] [Medline: 24526761]

Abbreviations

BTH: Blekinge Institute of Technology
GDS: Geriatric Depression Scale
MCI: mild cognitive impairment
MMSE: Mini-Mental State Exam

Original Paper

Thematic Analysis on User Reviews for Depression and Anxiety Chatbot Apps: Machine Learning Approach

Arfan Ahmed[1,2], PhD; Sarah Aziz[1,2*], MSc; Mohamed Khalifa[3*], PhD; Uzair Shah[1], BSc; Asma Hassan[1], MSc; Alaa Abd-Alrazaq[1,2], PhD; Mowafa Househ[1], PhD

[1]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

[2]AI Center for Precision Health, Weill Cornell Medicine-Qatar, Doha, Qatar

[3]Centre for Health Informatics, Australian Institute of Health Innovation, Faculty of Medicine, Health and Human Sciences, Macquarie University, Sydney, Australia

[*]these authors contributed equally

Corresponding Author:
Mowafa Househ, PhD
Division of Information and Computing Technology
College of Science and Engineering
Hamad Bin Khalifa University, Qatar Foundation
Liberal Arts and Sciences Building, Education City
Doha Al Luqta St, Ar-Rayyan
Doha, PO Box 5825
Qatar
Phone: 974 33223401
Email: mhouseh@hbku.edu.qa

Abstract

Background: Anxiety and depression are among the most commonly prevalent mental health disorders worldwide. Chatbot apps can play an important role in relieving anxiety and depression. Users' reviews of chatbot apps are considered an important source of data for exploring users' opinions and satisfaction.

Objective: This study aims to explore users' opinions, satisfaction, and attitudes toward anxiety and depression chatbot apps by conducting a thematic analysis of users' reviews of 11 anxiety and depression chatbot apps collected from the Google Play Store and Apple App Store. In addition, we propose a workflow to provide a methodological approach for future analysis of app review comments.

Methods: We analyzed 205,581 user review comments from chatbots designed for users with anxiety and depression symptoms. Using scraper tools and Google Play Scraper and App Store Scraper Python libraries, we extracted the text and metadata. The reviews were divided into positive and negative meta-themes based on users' rating per review. We analyzed the reviews using word frequencies of bigrams and words in pairs. A topic modeling technique, latent Dirichlet allocation, was applied to identify topics in the reviews and analyzed to detect themes and subthemes.

Results: Thematic analysis was conducted on 5 topics for each sentimental set. Reviews were categorized as positive or negative. For positive reviews, the main themes were confidence and affirmation building, adequate analysis, and consultation, caring as a friend, and ease of use. For negative reviews, the results revealed the following themes: usability issues, update issues, privacy, and noncreative conversations.

Conclusions: Using a machine learning approach, we were able to analyze ≥200,000 comments and categorize them into themes, allowing us to observe users' expectations effectively despite some negative factors. A methodological workflow is provided for the future analysis of review comments.

Introduction

Background

Mental health disorders can have a real impact on society, with ≥264 million people affected by depression alone, and anxiety accounts for an alarming 3.76% of the population globally [1]; depression is also a leading cause of disability [2]. When it comes to the working population, anxiety and depression are among the most commonly prevalent mental disorders worldwide, which can lead to high rates of sick leave and low job performance [3]. Smartphone-based chatbot apps with self-care interventions can provide cost-effective solutions compared with hospital-based therapeutic interventions. They can also increase the capacity of such care using assessment and mood tracking or by providing human-like conversation in times of loneliness. A chatbot is an artificial intelligence (AI) software that simulates conversations or chatting with users in natural language through messaging apps, webpages, and mobile apps or on the phone [3]. In a recent review, titled *A Review of Mobile Chatbot Apps for Anxiety and Depression and their self-care features*, published in Computer Methods and Programs in Biomedicine Update, we searched the Apple App Store and Google Play Store following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) protocol guidelines and identified 11 chatbot apps related to anxiety and depression, which were assessed for quality and characteristics [3]. Using known quality measures for mobile apps, such as authority, complementarity, confidentiality, validity, technological features, general information, usability, and intervention approaches, we found that available anxiety- and depression-related chatbot apps were highly rated and had high rates of user satisfaction [3]. The 11 apps included Ada, your health companion; Dr Sila, your smart health assistant; InnerHour Self-Care Therapy, anxiety and depression; MindDoc, depression and anxiety; Mindspa: Self Help 4UR Mental Wellbeing & Wellness; Pocketcoach, anxiety helper; Replika, My AI Friend; Serenity: Guided Mental Health; Woebot, your self-care expert in cognitive behavioral therapy and mindfulness; Wysa, stress, sleep, and mindfulness therapy chatbot; and Youper, anxiety and depression. These 11 apps targeted anxiety and depression disorders, and all contained chatbots. They were all of high quality and popularity and met the mHONCode principles [3]. The apps were built with various self-care interventions and other types of interventions, such as cognitive behavioral therapy and mindfulness. We found that anxiety and depression chatbot apps have the potential to improve the capacity of mental health self-care and provide low-cost support to professionals [3]. As a follow on from our initial study, we wanted to conduct in-depth analysis using a thematic review by extracting publicly available review comments and ratings from the Apple App Store and Google Play Store. By doing so, we gain better insight into the user satisfaction of these 11 apps. We considered this part 2 of our initial study, where we further explored users' opinions, satisfaction, and attitudes about anxiety and depression chatbot apps. Several studies have examined users' opinions and satisfaction with mental health apps but very few have analyzed app reviews in detail. For instance, a scoping review explored users' perceptions and opinions about

mental health chatbots, as reported by 37 previous studies using cross-sectional survey methods [4]. However, the scoping review, as well as the included 37 studies, did not analyze the qualitative reviews of the actual users of the apps, who downloaded and used the apps and then gave their feedback in app stores such as the Apple App Store and Google Play Store [4]. Web-based reviews reflect a user's perspective on a service or product. We consider these topics as dimensions along which different users evaluate the chatbots. Therefore, this study aims to explore users' opinions, satisfaction, and attitudes about anxiety and depression chatbot apps by conducting a thematic analysis of users' reviews of 11 anxiety and depression chatbot apps using topic modeling techniques. In addition, we propose a workflow to provide a methodological approach for future analysis of app review comments. Topic modeling is an unsupervised and automated machine learning (ML) technique that identifies abstract topics found in a body of documents using probabilistic models. They are frequently used as tools for text mining to find semantic structures in a text. For example, a text regarding certain topics will have specific words that appear more frequently than others. Such techniques are useful for classifying documents, organizing large chunks of texts, retrieving information from unstructured data, and selecting features. They are also more accurate and efficient than traditional data mapping approaches [5]. This review followed the thematic analysis approach, including data cleaning processes, clustering of reviews by one group of coauthors into positive or negative groups, validation of the accuracy of the clustering by a second group of coauthors, and finally analyzing the main topics extracted from each cluster and categorizing them into themes and subthemes. Investigating web-based reviews on apps designed for mental health conditions, such as depression and anxiety, is essential to understand users' preferences, feedback, and suggestions. This can add valuable insights into both clinical apps' development field and research in this rapidly evolving domain in relation to this specific clinical area.

Related Work

Previous studies of a similar nature have focused on various smartphone apps, and most of them successfully analyzed users' feedback and reviews. However, many studies failed to report their used methodologies in sufficient detail so that other research works could use such proposed approaches and build on what has been developed. A study in 2014 evaluated 229 medicine reminder apps, but their thematic analysis was limited to 1012 user reviews [5]. A study in 2017 analyzed users' emotions for ≥7 million reviews from the Apple App Store using a self-developed scraper tool. Such data scraping tools extract and import data from webpages and to spreadsheets to transform such data into human-readable output. They used a third-party sentiment extraction tool to extract meanings and impressions from short sentiment sets that include user views or opinions [6]. A recent study analyzed 63,398 reviews of e-commerce apps [7]. They conducted a thematic analysis and compared the performance of ML algorithms and Linguistic Inquiry and Word Count (LIWC). ML is an approach to data analysis in which the building of analytical models is automated. It is considered a part of AI, as it is based on the idea that systems can learn

from data, identify data patterns, and make decisions with minor human interventions [8]. Similarly, LIWC uses a predefined list of words in its dictionary to determine the positive and negative tones in English. The study reported a higher F1-score for the LIWC approach (86.7%), which statistically reflects a very high accuracy rate, based on the precision and recall of the test, where precision is known as the positive predictive value and recall is known as sensitivity in diagnostic binary classifications. A study in 2014 reported an average accuracy of 59% when using natural language processing (NLP) techniques to identify app features from reviews, followed by extracting user sentiments about these features and eventually used topic modeling techniques to group fine-grained features into more meaningful high-level features [9]. We found that various studies have developed tools to automate the analysis of reviews. One of them, a tool named Mining and Analyzing Reviews by Keywords, is a semiautomated tool to support the analysis of app reviews [10]. A study in 2018 was conducted as part of a 2-part study evaluating user experience through a thematic analysis of publicly available user reviews of cognitive behavioral therapy apps for depression [11]. They extracted 2904 reviews from 24 apps and manually assessed the reviews according to their sentiment (positive, negative, or neutral). Several studies [12] have used tools to allow thematic analysis, such as NVivo (QSR International), a qualitative analysis tool

[13]. The latter study [14] used a third-party tool to extract reviews and NLP for preprocessing the data and developed several ML classifiers to predict the polarity (positive or negative) of reviews. They trained the classifiers for prediction and then used NVivo to conduct thematic analysis. We reviewed existing studies and found no clear consistent approach to conduct a similar analysis. Therefore, based on existing literature, we propose our own workflow, which includes an ML technique to automatically analyze text data (topic modeling), and aim to also provide a methodological approach for future analysis of review comments.

Methods

Overview

We aimed to identify how 11 depression and anxiety chatbots were scrutinized by users, as shown in Table 1.

Using the Apple App Store and Google Play Store, we extracted 205,581 reviews and classified them under positive and negative headings and used the following workflow. There were six steps, as illustrated in Figure 1: data collection, data preprocessing, data annotation, data vectorization, topic modeling, and thematic analysis. Further details regarding these steps are presented in the following subsections.

Table 1. Apps, supported platforms, and total reviews.

App name	Platform	Reviews (N=205,581), n (%)
Ada: your health companion	Android	60,441 (29.4)
Dr Sila: your smart health assistant	Android and iOS	2 (<0.01)
InnerHour Self-Care Therapy: anxiety and depression	Android and iOS	1975 (0.96)
MindDoc: depression and anxiety	Android and iOS	8010 (3.9)
Mindspa: Self Help 4UR Mental Wellbeing & Wellness	Android and iOS	43 (0.02)
Pocketcoach: anxiety helper	Android and iOS	83 (0.04)
Replika: My AI Friend	Android and iOS	102,534 (49.88)
Serenity: Guided Mental Health	Android and iOS	3328 (1.62)
Woebot: your self-care expert in CBT[a] and mindfulness	Android and iOS	4316 (2.1)
Wysa: stress, sleep, and mindfulness therapy chatbot	Android and iOS	24,349 (11.84)
Youper: anxiety and depression	iOS	500 (0.24)

[a]CBT: cognitive behavioral therapy.

Figure 1. Workflow to prepare reviews for thematic analysis.
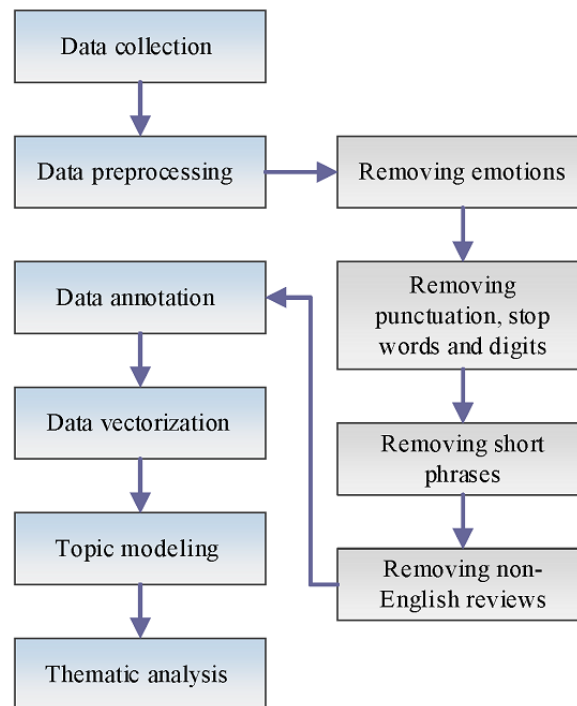
Data Collection

We used 11 apps from our previous study, *A Review of Mobile Chatbot Apps for Anxiety and Depression and their self-care features*, published in Computer Methods and Programs in Biomedicine Update [3]. In this previous study, a search was performed in the Apple App Store and Google Play Store following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) protocol to identify existing chatbot apps for anxiety and depression. The eligibility of the studies was assessed by two researchers (AA and SA), based on 5 predefined eligibility criteria. Eligibility criteria outlined that the app should be related to either anxiety or depression or both, the apps should contain a chatbot feature as opposed to human interaction as the main chatting agent, the apps should be free of cost at the point of download, the apps that have ratings above 4 stars as a sign of user satisfaction only should be included, and only apps that had a total number of ≥5 raters were included [3]. Metadata of the included chatbots and their characteristics were extracted from their description and after installation by two reviewers (AA and SA). We extracted 205,581 user reviews from the 11 apps identified in our previous study using a scraper tool that we built using the Google Play Scraper and Apple App Store Scraper Python libraries. All Google Play Store reviews for the 11 apps were extracted, whereas for the Apple App Store, it only allowed a maximum of 500 reviews per app to be extracted because of privacy policies. Table 1 outlines the selected apps and the total number of consumer reviews extracted from each app.

Data Preprocessing

Web-based reviews can contain many irrelevant comments, such as those that do not describe the apps or mention their advantages and disadvantages. Therefore, preprocessing the reviews using NLP techniques was needed in the following stages:

1. Normalizing the data, converting the text of all reviews to lowercase.
2. Removing emoticons, symbols and pictographs, and transport and map symbols from the reviews.
3. Removing punctuation marks and digits from the reviews using the natural language toolkit Python library.
4. Removing short phrases (eg, "Good," "Nice," "very Helpful," or "really impressive app"). As we are identifying the key themes or features or issues pointed out by reviewers, short phrases provide no additional value to our thematic analysis.
5. Removing non-English reviews (eg, German and Dutch as app secondary languages).

Data Annotation

A star rating score on a scale of 1 to 5 was assigned by app users within each user review, which was part of our extracted data. One star represents *very dissatisfied* and 5 stars represents *very satisfied*. The reviews were divided into 2 sets by annotating them as either positive or negative sentiment polarity based on the rating given to each review. We adopted this approach from previous studies that categorized the data in a similar manner [15]. Table 2 outlines the criteria adopted for mapping sentiment polarities to user ratings.

Neutral feedback was considered not positive and was included in the negative sentiment polarity. We further validated our technique for accuracy by randomly selecting 50 positive and negative reviews, 2 reviewers manually validated the results, and our results showed 100% accuracy for annotation on the 50 respective sentiment reviews.

Table 2. Criteria for user rating to polarity mapping.

User ratings	Rating description	Sentiment polarity
1	Very dissatisfied	Negative
2	Dissatisfied	Negative
3	Neutral	Negative
4	Satisfied	Positive
5	Very satisfied	Positive

Data Vectorization

For the model to handle the data, it must be represented in numerical form. One of the most popular ways to achieve this is the word's Term Frequency–Inverse Document Frequency score [16] or their frequency counts (bag-of-words [BOW] approach). We opted for the BOW approach [17] representation to vectorize the terms by identifying the unique terms in our corpus and used the *CounterVectorizer* function from *Sklearn's* feature extraction module in Python. This function converts a collection of text to a matrix of word counts, allowing us to extract bigrams and use a count vectorizer, which is a known technique for retrieving the relevant counts of each bigram.

Topic Modeling

Topic modeling is based on unsupervised and automated ML techniques that discover abstract topics in the body of text using probabilistic models. These techniques are frequently used as text mining tools to identify semantic structures and conduct thematic analyses [18]. It is considered superior to other manual thematic analysis techniques, traditional statistical models, or standard text mining techniques. Manual thematic analysis techniques [14] applied with software tools such as NVivo use themes or topics that are manually predefined by the reviewer, and the corresponding reviews are further mapped, which can be challenging when the amount of data is large. Traditional statistical techniques are used for clustering data, primarily K-means clustering, which is a method that partitions data observations into data clusters based on the nearest mean. This method works on numeric data as opposed to text data. Standard text mining techniques use keyword information representation and count-based mining of these keywords instead of analyzing the use of these keywords and their contexts. Topic modeling has an advantage over the aforementioned techniques in that it performs clustering on the textual data provided inside of the data in depth. Furthermore, it is useful for identifying the key issues and topics discussed in large corpora. Topic modeling algorithms use latent Dirichlet allocation (LDA), which determines a word's meaning based on its co-occurrence with other words. In this study, we used a bigram LDA model for topic modeling. LDA is a Bayesian inference method that estimates the most likely topics given for the observed corpus.

We used the LDA model and analyzed the bigram range of the corpus of reviews in each set.

Furthermore, we divided our data set of reviews into 2 sets of positive and negative reviews and performed topic modeling on each set separately to identify the main themes within our data set. We worked with 5 topics from each set of sentimental data, selected through consensus after deeming the results of 3 topics as inappropriate.

Thematic Analysis

Thematic analysis was conducted on the 5 topics of each sentimental set. Through these 5 topics of negative and positive sets, we were able to identify the factors that contribute to the effectiveness of anxiety and depression chatbots both negatively and positively. We identified the topic names and analyzed them thoroughly to identify the themes encapsulated within each theme.

Ethics Approval

No ethics approval needed, as we used publicly available data.

Results

Data Collection, Preprocessing, Annotation, and Vectorization

We analyzed 205,581 comments extracted from 11 anxiety and depression chatbots identified in our previously published study on both the Apple App Store and Google Play Store [3]. The steps of review selection and inclusion are shown in Figure 2. Of the 205,581 reviews, 1093 (0.53%) reviews were removed, containing only emoticons (smiley face, sad face, heart shape, etc). From the remaining 204,488 reviews containing only text, a further 1360 (0.67%) non-English reviews were excluded. Furthermore, 32.48% (66,423/204,488) reviews with simple praises were removed, as they did not contribute anything meaningful to our analysis, leaving a remainder of 136,705 reviews.

The 136,705 reviews were divided into two meta-themes: positive and negative, as presented in Table 3.

Multimedia Appendix 1 shows the complete app-by-app breakdown of the preprocessing and meta-theme division.

Figure 2. Steps for the selection and inclusion of reviews.
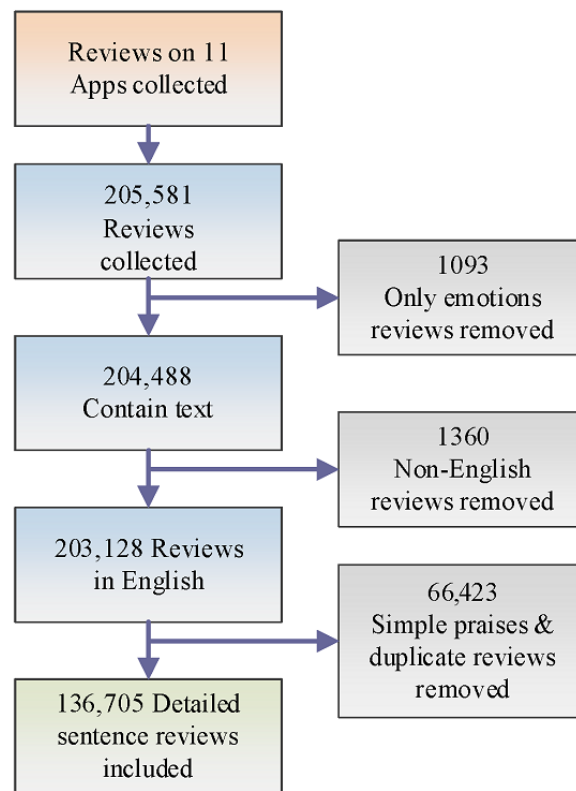
Table 3. Sentiment polarity with corresponding reviews.

Sentiment polarity	Number of reviews
Positive	124,458
Negative	12,247

Topic Modeling and Thematic Analysis

Topic Extraction Model for Positive and Negative Reviews

Having preprocessed (cleaned) the data, we divided it into two meta-themes (positive and negative). Topic modeling was performed on each set of reviews separately, as per previous studies, as it is said to be a fast approach for identifying themes in large corpora [19]. We identified 5 topics from the positive and negative reviews, forming 5 topics in each set. Multimedia Appendix 2 presents the most common words used in each set of topics. The relative composition in the corpus with respect to each topic was calculated using LDA by computing the appearance of each topic occurring with respect to the underlining review. Thus, each review obtained the appearance frequencies corresponding to each topic, and the highest value indicated the review mapping into that respective topic. Each topic was assigned a name, and the respective subthemes were identified by analyzing the number of reviews in each respective topic. Multimedia Appendix 3 depicts the respective topics in each set of reviews.

Categorizing Themes

All subthemes identified in each topic were classified under one specific larger theme. Although the topics addressed different issues, identifying the main themes allowed them to be grouped together under the same category and further analyzed. More than 90% of the reviews were classified as positive, and <10% were classified as negative. Figure 3 shows the themes and subthemes from both positive and negative perspectives. Figure 3 shows the meta-theme categories for positive and negative reviews, along with each theme-encapsulated subtheme. For positive reviews, the main themes were "confidence and affirmation building," "adequate analysis and consultation," "caring as a friend," and "easy to use." For negative reviews, the main themes were "usability issues," "update issues," "privacy," and "noncreative conversation."

Figure 3. Meta-themes, themes, and subthemes for positive and negative reviews.
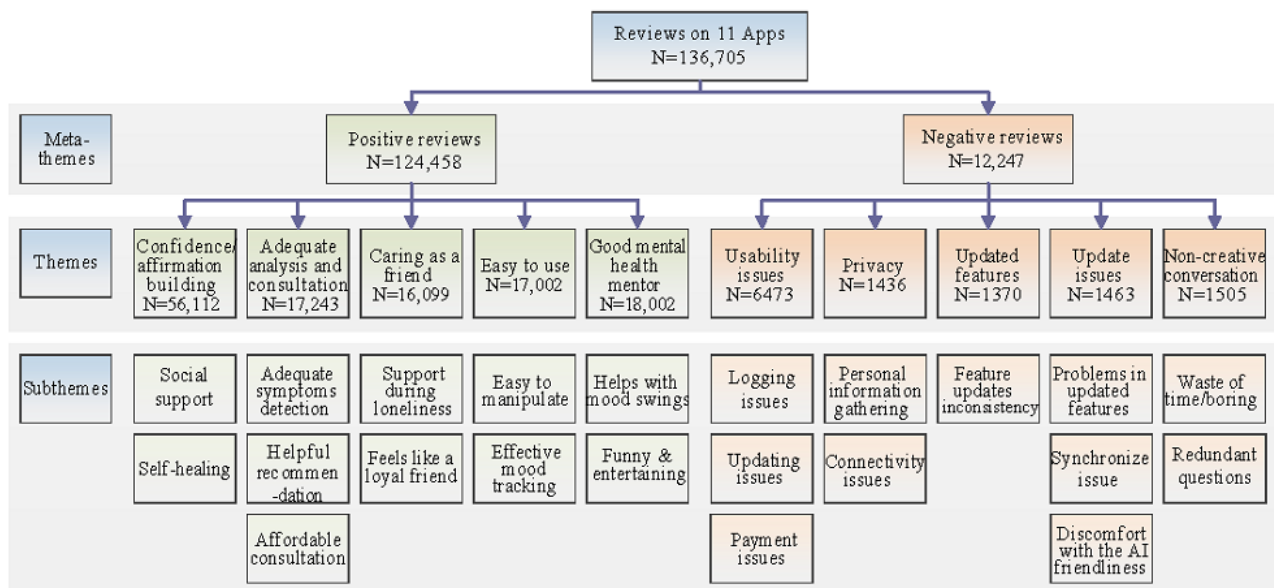
Positive Reviews

The following section highlights the positive experiences people reported while using the 11 anxiety and depression chatbots. A sample of positive reviews is included in Multimedia Appendix 4 and classified into 4 themes.

Theme 1: Confidence and Affirmation Building

The first theme, confidence and affirmation building, consists of three subthemes: (1) social support, (2) self-healing, and (3) help with mood swings, as illustrated in Figure 1. Social support was a major concern for users, the chatbot provides comfort by appreciating even the smallest of users' achievements and providing encouragement, allowing users to believe in themselves and overcoming social anxieties. The second subtheme shows the ability of the chatbot to provide users with tools to overcome self-control issues, allowing users to provide their own solution and raising self-esteem and will power, as well as allowing them to feel comfortable in their own body. The third subtheme states that chatbots allow the control and tracking of mood swings for better treatment.

Theme 2: Adequate Analysis and Consultation

The second theme consists of three subthemes: (1) adequate symptom detection, (2) helpful recommendations, and (3) affordable consultation. Adequate symptom detection was considered an important aspect for chatbot users selected in this review. The chatbots can relate to users' symptoms and provide empathy. The following subtheme is *helpful recommendation* in which chatbots support recommendations that are synchronized with the user's counselor or health care professional. Affordable consultation is the third subtheme in which users highlight that they appreciate the effectiveness of the chatbot with the bonus of using a consultation app with minimal to no cost.

Theme 3: Caring as a Friend

The third theme is composed of three subthemes: (1) support during loneliness, (2) feeling like a loyal friend, and (3)

entertaining talks. The first subtheme involves chatbots becoming the user's go-to person during times of loneliness. The second subtheme is identified as *feeling like a loyal friend* because users indicated that having someone to talk to is not always the only solution; trusting that person as a friend is also important. Users felt that chatbots could provide this, as reflected in their comments. The third subtheme highlights that engaging in conversations containing humor provides a boost to users' mood and is entertaining. Slang language and memes are also enjoyed by users. The chatbots had the intelligence to understand the mood of a person and to boost their mood accordingly.

Theme 4: Easy to Use

The last theme in the positive reviews includes two subthemes: (1) easy to manipulate and (2) effective mood tracking. The first subtheme is easy to manipulate. By manipulation, we refer to how conversations with chatbots occur in an easy manner in which the questions are easy to be followed and understood. In addition, chatbots are easy to use and require minimal assistance. The anxiety and depression chatbots were easy to navigate, which the users appreciated. The second subtheme is effective mood tracking, in which the chatbot assigns tasks to individuals that are easily achievable with good results.

Negative Reviews

The following section describes users' concerns regarding the negative experiences encountered while using chatbots. A sample of negative reviews is included in Multimedia Appendix 4 and classified into 4 themes.

Theme 1: Usability Issues

This theme outlines various usability issues that are encountered while accessing the different functionality of the chatbots. Subthemes identifying these issues are (1) logging issues, (2) payment issues, and (3) connectivity issues. Nonresponsive interfaces or lack of guidance to navigate with regard to signing is one of the major issues encountered. The second subtheme highlights the issue regarding payments; although most apps

were free, people did not take well to in-app purchases in some cases, and these were reflected in some comments. The last subtheme signifies the connectivity glitches experienced by many while using the apps. Although the apps connected well via the internet, some server accessibility issues were observed, which left a negative taste for users.

Theme 2: Update Issues

The second theme, updating the apps, consists of two subthemes: (1) problems in updated features and (2) synchronization issues. The updates caused people to drop out of the apps by interrupting their ongoing processes or tasks. Changes in features or added costs to previously free features post updates were not welcomed by many users. The second subtheme states a lack of synchronization between different app versions. The version updates made simple tasks more difficult in newer versions or were sometimes removed completely without considering current users.

Theme 3: Privacy

The third theme, privacy, consists of two subthemes: (1) personal information gathering and (2) discomfort with AI friendliness. The first subtheme talks about users' dislike when many personal questions are asked. The second subtheme reflects the reviews emphasizing the discomfort they experience while talking to the chatbots; users expressed that some dialog used by the chatbots was inappropriate, as indicated in sample reviews, making users feel they are under surveillance or being interrogated, and some highlighting how the AI bot becomes too friendly.

Theme 4: Noncreative Conversation

The first theme, confidence and affirmation building, consists of two subthemes: (1) waste of time or boring and (2) redundancy of questions. The first subtheme was developed as users continuously emphasized that the chatbot is wasting their time and it is becoming boring. Short-term memory of the chatbot and not building upon previous conversations was also highlighted, along with chats turning out to be monotonous as the questions were repeated with no new engaging questions asked. The second subtheme is the redundancy of questions, in which users emphasize that the questions are repetitive questions daily, and some sentence structure change is required.

Discussion

Principal Findings

Chatbots are promising tools for those who have anxiety and depression, providing confidence and support via user-friendly tools. Users enjoy engaging and creative content while remaining conscious of security issues. Through thematic analysis, 4 positive and 4 negative themes and relevant 12 subthemes in each group were identified in the users' reviews of 11 depression and anxiety chatbot apps. Positive reviews highlighted the features that users enjoyed most using chatbots. Users identified several areas in which we clustered into two major parts: positive and negative reviews. Our methodological approach identified thematic content by categorizing a large corpus of reviews into 5 clusters based on each sentimental division set. Important inferences can be deduced from the

empirical findings, and most of the reviews were positive, indicating that the majority of users found anxiety and depression chatbots effective and fit for purpose. Negative experiences are also encountered while using these chatbots.

In Figure 3, our findings revealed that the majority of the 11 anxiety and depression chatbots worked well in this regard, boosting a user's social support, self-healing, and dealing with mood swings. The first theme is confidence and affirmation building, which is a key aspect that allows a person to maintain their well-being, especially when it comes to mental health. Affirmation generally works as a tool for mindset shifting and achieving goals through social support [20]. Finally, as chatbots help patients with mood swings, it is important to highlight that mood variability is a major trait in those affected by anxiety and depression [16]. The second theme in positive reviews was adequate analysis and consultation. Most health apps that are currently available lack clinically validated evidence of their efficacy [21]. Our findings revealed that most of these anxiety and depression chatbots played their part in the best possible way by providing symptom detection and helpful recommendations with the best affordable consultation one can expect with freely available apps.

This is aligned with the findings of a previous study [22], which highlighted the issue of waiting times and the affordability of psychotherapy. Chatbots can provide access to affordable care before waiting for approval to meet a clinician [23]. The third theme is *caring as a friend*, and the evidence behind it is that feelings of loneliness are a major factor in mental health issues [21]. Loneliness and social isolation are tantamount to feeling unsafe and are classified as social threats accompanied by feelings of hostility, low self-esteem, anxiety, stress, and pessimism [24]. This issue has been exacerbated during the COVID-19 pandemic, resulting in a large number of mental health issues [25]. Many of the review comments reveal how users benefited from chatbots during dark times, especially during the current pandemic.

Issues of accessibility were highlighted when using apps. An app has to be *easy to use* [26] and must be a key design thought when developing mental health–related apps. Users are less accepting of complex scenarios and tasks, particularly when they are already affected. We observed that users regarded the chatbots as easily operable, and the tasks were quick and to the point.

The first observed theme among negative reviews was related to usability issues. Understanding user experience is essential for chatbot design to overcome usability issues that could be achieved by applying usability testing metrics such as the system usability scale and user experience questionnaire scales to measure and evaluate user experience and satisfaction [27].

From the perspective of usability, effectiveness, efficiency, and satisfaction [28] are key considerations when achieving a specific task or goal. We found that users disliked issues related to logging or startups. The last 2 subthemes in this section are related to payment and connectivity issues, the cost of apps, and problems with connectivity, which are powerful platforms that do not rely on costly servers [29]. The second theme in the negative reviews highlighted problems related to new updates

in the app that caused the chatbot to lose its essence or meaning, as some users did not like the inconsistency in updates. Although updates are welcomed, users become frustrated whenever an update of previously running features is blocked or they are asked to pay for them.

Privacy is the third theme in negative reviews, which is a key issue that users face with chatbots [30]. The subthemes are divided into personal information gathering–related issues and observations with users who do not feel comfortable with the overfriendliness of the AI robotic chatbot. Users become wary of what and with whom they share personal conversations, and many ethical and legal (data protection) questions remain unanswered [31]. In addition, in our subtheme, discomfort with AI friendliness was in line with the findings of previous studies [32], showing that users linked the technical systems with human-like attributes, which makes users feel uncomfortable about content sharing and possible data sharing and general mistrust toward the app. The last theme in the negative reviews is noncreative conversation. Unsurprisingly, unintelligent conversations proved a big dislike to users. The creativity of the chatbot is questioned when a chatbot is repetitive, asking similar questions throughout the day, and users lose interest. Questions also arise about the susceptibility of users to receive therapeutic advice from chatbots, whose algorithms may be harmful to users [31]. The frequency of themes, shown in Multimedia Appendix 3 and Figure 3, should help app developers understand the main advantages that users highlight in their reviews and the main challenges they faced. Consequently, developers should work to fix the weaknesses of their apps and invest more in areas that users consider important to manage their medical conditions. Eventually, this should improve the public adoption of apps designed for mental health conditions such as anxiety and depression.

Strengths and Limitations

This study has several strengths and limitations regarding the review analysis of anxiety and depression chatbots. The study enabled us to explore such a large review corpus in a short span, which was otherwise not possible. To the best of our knowledge, no such thematic analysis is available using an ML approach. This approach provides a high-level view of all the possible themes that the topics encapsulate. However, this study was analyzed on English reviews only. In addition, the Apple App Store privacy policy did not allow researchers to exceed 500 review extractions per app; thus, the findings may be restricted in view of Apple users of the app. Moreover, reviews that sometimes included emojis or emoticons only were excluded from this thematic analysis. The study method approach of combining neutral and negative into one category could be seen as a limitation, although this methodology has been reported in previous studies. Finally, our findings are limited to 11 chatbots related to anxiety and depression, and we emphasize that the generalizability and applicability of our results may not be

applicable to other chatbots in the market outside of mental health.

Practical and Research Implications

Practical Implications

This review serves most researchers, clinical app developers, and professionals interested in developing chatbots related to mental health and well-being. We identify a gap in research as researchers highlight issues related to clinical psychotherapy approaches and neglect the psychosocial aspect of using mental health apps. We highlight factors related to the social, technical, and understanding of user experiences and opinions. On the basis of positive and negative reviews, users were put off by issues related to privacy, repetitiveness of conversation by the chatbot, connectivity issues, and poor usability. Among the positive themes, we found that ease of use, providing support, good human conversation, and providing feelings of mood uplifting were perceived as indicators of desired attributes. With the involvement of health care providers and policy makers, future development needs to seriously consider our findings when developing such chatbots by chatbot app developers and researchers.

Research Implications

Analysis of user reviews has been explored previously, and although powerful techniques exist, we found that clear guidance was lacking when we looked at previous studies on how to implement such techniques for large amounts of review data. In this study, we documented our approach and invited other researchers to test and validate our proposed method and to check its feasibility in analyzing users' reviews of other apps. In addition, future developers of mental health–related chatbots can use our findings to build chatbots with more positive opinions and higher rates of satisfaction.

Conclusions

Through thematic analysis of users' reviews, this study explored opinions, satisfaction, and attitudes regarding 11 anxiety and depression chatbot apps. Our proposed workflow provides a methodological approach for the future analysis of app review comments. We used an ML-based technique known as topic modeling to cluster reviews to obtain insight into textual data, as opposed to traditional data-mapped approaches that determine themes. Users tend to dislike technical and privacy issues. Users expect engaging and creative conversations through more appealing user interfaces. Moreover, users of chatbots designed for anxiety and depression feel supported and confident as they use apps that are easy to manipulate, affordable, and free of cost. Consequently, users do not prefer chatbots with less creative content and conversations and those that are overly friendly and invade their privacy. These themes allowed us to observe users' expectations effectively, despite some negative factors. Future researchers can work toward better generalizing topics by closely refining reviews into specific issues or features.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1
Complete app-by-app breakdown of the preprocessing and meta-theme division.
[DOCX File , 23 KB - formative_v6i3e27654_app1.docx]

Multimedia Appendix 2
The most common words used in each set of topics.
[DOCX File , 22 KB - formative_v6i3e27654_app2.docx]

Multimedia Appendix 3
Respective topics in each set of reviews.
[DOCX File , 25 KB - formative_v6i3e27654_app3.docx]

Multimedia Appendix 4
Examples of comments.
[DOCX File , 49 KB - formative_v6i3e27654_app4.docx]

References

1. Elflein J. Share of the population worldwide who suffered from anxiety disorders from 1990 to 2017. 2019 2020. Statista. 1990. URL: https://www.statista.com/statistics/1035153/percentage-of-people-with-anxiety-worldwide/ [accessed 2020-12-30]
2. Rudd BN, Beidas RS. Digital mental health: the answer to the global mental health crisis? JMIR Ment Health 2020 Jun 02;7(6):e18472 [FREE Full text] [doi: 10.2196/18472] [Medline: 32484445]
3. Brenninkmeijer V, Lagerveld SE, Blonk RW, Schaufeli WB, Wijngaards-de Meij LD. Predicting the effectiveness of work-focused CBT for common mental disorders: the influence of baseline self-efficacy, depression and anxiety. J Occup Rehabil 2019 Mar 15;29(1):31-41. [doi: 10.1007/s10926-018-9760-3] [Medline: 29450678]
4. Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M. Perceptions and opinions of patients about mental health chatbots: scoping review. J Med Internet Res 2021 Jan 13;23(1):e17828 [FREE Full text] [doi: 10.2196/17828] [Medline: 33439133]
5. Storz MA. To the Editor:. Menopause 2020 Nov;27(11):1343. [doi: 10.1097/GME.0000000000001617] [Medline: 33110053]
6. Martens D, Johann T. On the emotion of users in app reviews. In: Proceedings of the IEEE/ACM 2nd International Workshop on Emotion Awareness in Software Engineering (SEmotion). 2017 May 21 Presented at: IEEE/ACM 2nd International Workshop on Emotion Awareness in Software Engineering (SEmotion); May 21-21, 2017; Buenos Aires, Argentina. [doi: 10.1109/semotion.2017.6]
7. Olagunju T, Oyebode O, Orji R. Exploring key issues affecting african mobile ecommerce applications using sentiment and thematic analysis. IEEE Access 2020;8:114475-114486. [doi: 10.1109/access.2020.3000093]
8. Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. Electr Market 2021 Apr 08;31(3):685-695. [doi: 10.1007/s12525-021-00475-2]
9. Guzman E, Maalej W. How do users like this feature? A fine grained sentiment analysis of app reviews. In: Proceedings of the IEEE 22nd International Requirements Engineering Conference (RE). 2014 Aug 25 Presented at: IEEE 22nd International Requirements Engineering Conference (RE); Aug. 25-29, 2014; Karlskrona, Sweden. [doi: 10.1109/RE.2014.6912257]
10. Vu P, Pham H, Nguyen T, Nguyen T. Tool support for analyzing mobile app reviews. In: Proceedings of the 30th IEEE/ACM International Conference on Automated Software Engineering (ASE). 2015 Nov 09 Presented at: 30th IEEE/ACM International Conference on Automated Software Engineering (ASE); Nov. 9-13, 2015; Lincoln, NE, USA. [doi: 10.1109/ase.2015.101]
11. Stawarz K, Preist C, Tallon D, Wiles N, Coyle D. User experience of cognitive behavioral therapy apps for depression: an analysis of app functionality and user reviews. J Med Internet Res 2018 Jun 06;20(6):e10120 [FREE Full text] [doi: 10.2196/10120] [Medline: 29875087]
12. Milward J, Khadjesari Z, Fincham-Campbell S, Deluca P, Watson R, Drummond C. User preferences for content, features, and style for an app to reduce harmful drinking in young adults: analysis of user feedback in app stores and focus group interviews. JMIR Mhealth Uhealth 2016 May 24;4(2):e47 [FREE Full text] [doi: 10.2196/mhealth.5242] [Medline: 27220371]

13. NVivo: Qualitative Data Analysis Software. QS International. 2020. URL: https://www.qsrinternational.com/nvivoqualitative-data-analysis-software/home [accessed 2020-04-30]

14. Oyebode O, Alqahtani F, Orji R. Using machine learning and thematic analysis methods to evaluate mental health apps based on user reviews. IEEE Access 2020;8:111141-111158. [doi: 10.1109/access.2020.3002176]

15. Gebauer J, Tang Y, Baimai C. User requirements of mobile technology: results from a content analysis of user reviews. Inf Syst E-Bus Manage 2007 Dec 7;6(4):361-384. [doi: 10.1007/s10257-007-0074-9]

16. Bowen R, Baetz M, Hawkes J, Bowen A. Mood variability in anxiety disorders. J Affect Disord 2006 Apr;91(2-3):165-170. [doi: 10.1016/j.jad.2005.12.050] [Medline: 16458367]

17. Zhang Y, Jin R, Zhou Z. Understanding bag-of-words model: a statistical framework. Int J Mach Learn & Cyber 2010 Aug 28;1(1-4):43-52. [doi: 10.1007/s13042-010-0001-0]

18. Wallach H. Topic modeling: beyond bag-of-words. In: Proceedings of the 23rd international conference on Machine learning. 2006 Jun 25 Presented at: ICML '06: Proceedings of the 23rd international conference on Machine learning; June 25 - 29, 2006; Pittsburgh Pennsylvania USA. [doi: https://doi.org/10.1145/1143844.1143967]

19. Piepenbrink A, Gaur AS. Topic models as a novel approach to identify themes in content analysis. Academy of Management 2017 Aug;2017(1):11335. [doi: 10.5465/AMBPP.2017.141]

20. White MA. Positive affirmations: too good to be true? Healthline. 2020 Sep 01. URL: https://www.healthline.com/health/mental-health/do-affirmations-work [accessed 2021-01-01]

21. Richardson T, Elliott P, Roberts R. Relationship between loneliness and mental health in students. J Public Ment Health 2017 Jun 19;16(2):48-54. [doi: 10.1108/jpmh-03-2016-0013]

22. Grünzig SD, Baumeister H, Bengel J, Ebert D, Krämer L. Effectiveness and acceptance of a web-based depression intervention during waiting time for outpatient psychotherapy: study protocol for a randomized controlled trial. Trials 2018 May 22;19(1):1-11 [FREE Full text] [doi: 10.1186/s13063-018-2657-9] [Medline: 29788996]

23. Denecke K, Abd-Alrazaq A, Househ M. Artificial intelligence for chatbots in mental health: Opportunities and challenges. In: Multiple Perspectives on Artificial Intelligence in Healthcare. Cham: Springer; 2021:115-128.

24. Hawkley LC, Cacioppo JT. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. Ann Behav Med 2010 Oct 22;40(2):218-227 [FREE Full text] [doi: 10.1007/s12160-010-9210-8] [Medline: 20652462]

25. Lee CM, Cadigan JM, Rhew IC. Increases in Loneliness Among Young Adults During the COVID-19 Pandemic and Association With Increases in Mental Health Problems. J Adolesc Health 2020 Nov;67(5):714-717 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.08.009] [Medline: 33099414]

26. Dragovic M, Davison S, Morgan VA, Chiu VW, Richards N, Vatskalis T, et al. 'Validated, easy to use and free': top three requests for mobile device applications ('apps') from mental health consumers and clinicians. Adv Ment Health 2018 Dec 11;18(2):106-114. [doi: 10.1080/18387357.2018.1557014]

27. Holmes S, Moorhead A, Bond R, Zheng H, Coates V, McTear M. Usability testing of a healthcare chatbot: can we use conventional methods to assess conversational user interfaces? In: Proceedings of the 31st European Conference on Cognitive Ergonomics. 2019 Presented at: ECCE 2019: 31st European Conference on Cognitive Ergonomics; September 10 - 13, 2019; Belfast, Northern Ireland, UK p. 207-214. [doi: 10.1145/3335082.3335094]

28. Jokela T, Iivari N, Matero J, Karukka M. The standard of user-centered design and the standard definition of usability: analyzing ISO 13407 against ISO 9241-11. In: Proceedings of the Latin American conference on Human-computer interaction. 2003 Presented at: CLIHC03: Latin American Conference on Human-Computer Interaction; August 17 - 20, 2003; Rio de Janeiro Brazil p. 53-60. [doi: 10.1145/944519.944525]

29. Kirby R. Get ready for the genomic revolution. Trends Urol Men Health 2017 Apr 10;8(2):4. [doi: 10.1002/tre.567]

30. Ischen C, Araujo T, Voorveld H, van Noort G, Smit E. Privacy concerns in chatbot interactions. In: Chatbot Research and Design. 2020 Presented at: International Workshop on Chatbot Research and Design; November 19-20, 2019; Amsterdam, The Netherlands p. 34-48. [doi: 10.1007/978-3-030-39540-7_3]

31. Bendig E, Erb B, Schulze-Thuesing L, Baumeister H. The next generation: chatbots in clinical psychology and psychotherapy to foster mental health – a scoping review. Verhaltenstherapie 2019 Aug 20:1-13. [doi: 10.1159/000501812]

32. Erel H, Trayman D, Levy C, Manor A, Mikulincer M, Zuckerman O. Can you tell the difference? Comparing face-to-face versus computer-based interventions. The "Eliza" effect in psychotherapy. J Cogn Behav Psychother 2013;13(2):291-298 [FREE Full text]

Abbreviations

AI: artificial intelligence
LDA: latent Dirichlet allocation
LIWC: Linguistic Inquiry and Word Count
ML: machine learning
NLP: natural language processing
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

XSL•FO
RenderX

Original Paper

Implementing Machine Learning Models for Suicide Risk Prediction in Clinical Practice: Focus Group Study With Hospital Providers

Kate H Bentley[1,2,3], PhD; Kelly L Zuromski[2], PhD; Rebecca G Fortgang[2], PhD; Emily M Madsen[1,4], BSc; Daniel Kessler[2], MA; Hyunjoon Lee[1,4], MSc; Matthew K Nock[2], PhD; Ben Y Reis[3,5], PhD; Victor M Castro[6], MSc; Jordan W Smoller[1,3,4], PhD

[1]Center for Precision Psychiatry, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States

[2]Department of Psychology, Harvard University, Cambridge, MA, United States

[3]Harvard Medical School, Boston, MA, United States

[4]Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, United States

[5]Predictive Medicine Group, Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, United States

[6]Research Information Science and Computing, Mass General Brigham, Somerville, MA, United States

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

Abstract

Background: Interest in developing machine learning models that use electronic health record data to predict patients' risk of suicidal behavior has recently proliferated. However, whether and how such models might be implemented and useful in clinical practice remain unknown. To ultimately make automated suicide risk–prediction models useful in practice, and thus better prevent patient suicides, it is critical to partner with key stakeholders, including the frontline providers who will be using such tools, at each stage of the implementation process.

Objective: The aim of this focus group study is to inform ongoing and future efforts to deploy suicide risk–prediction models in clinical practice. The specific goals are to better understand hospital providers' current practices for assessing and managing suicide risk; determine providers' perspectives on using automated suicide risk–prediction models in practice; and identify barriers, facilitators, recommendations, and factors to consider.

Methods: We conducted 10 two-hour focus groups with a total of 40 providers from psychiatry, internal medicine and primary care, emergency medicine, and obstetrics and gynecology departments within an urban academic medical center. Audio recordings of open-ended group discussions were transcribed and coded for relevant and recurrent themes by 2 independent study staff members. All coded text was reviewed and discrepancies were resolved in consensus meetings with doctoral-level staff.

Results: Although most providers reported using standardized suicide risk assessment tools in their clinical practices, existing tools were commonly described as unhelpful and providers indicated dissatisfaction with current suicide risk assessment methods. Overall, providers' general attitudes toward the practical use of automated suicide risk–prediction models and corresponding clinical decision support tools were positive. Providers were especially interested in the potential to identify high-risk patients who might be missed by traditional screening methods. Some expressed skepticism about the potential usefulness of these models in routine care; specific barriers included concerns about liability, alert fatigue, and increased demand on the health care system. Key facilitators included presenting specific patient-level features contributing to risk scores, emphasizing changes in risk over time, and developing systematic clinical workflows and provider training. Participants also recommended considering risk-prediction windows, timing of alerts, who will have access to model predictions, and variability across treatment settings.

Conclusions: Providers were dissatisfied with current suicide risk assessment methods and were open to the use of a machine learning–based risk-prediction system to inform clinical decision-making. They also raised multiple concerns about potential

XSL·FO

RenderX

barriers to the usefulness of this approach and suggested several possible facilitators. Future efforts in this area will benefit from incorporating systematic qualitative feedback from providers, patients, administrators, and payers on the use of these new approaches in routine care, especially given the complex, sensitive, and unfortunately still stigmatized nature of suicide risk.

Introduction

Background

It is estimated that approximately 800,000 people die by suicide each year worldwide, representing approximately 1 suicide every 40 seconds [1]. In the United States, suicide is the 10th leading cause of death [2]. More than 48,000 Americans die by suicide each year, which works out to approximately 129 suicide deaths every day [2]. Encouragingly, from 2018 to 2019, the US suicide rate declined (by 2.1%) for the first time after 13 years of consecutive increases [2]. However, most states did not experience significant decreases, and whether this downward trend will continue remains to be determined [3]. To achieve the ambitious goal of reducing the suicide rate by 20% before by 2025 in the United States [4], it is critical to prioritize the large-scale implementation of existing evidence-based suicide prevention strategies, as well as the development of new approaches.

The health care system is a key setting in which to target suicide risk detection and prevention efforts. Estimates show that upwards of 75% of people who die by suicide pass through the health care system within the year (and up to 50% within the month) before their death [5]. Visits in medical specialty and primary care settings, followed by the emergency department (ED), are the most common leading up to death by suicide; notably, most of these visits do not include a documented mental health diagnosis [6]. In part owing to these findings, increased national efforts have been made to implement standardized self-reported or clinician-rated tools to assess suicidal thoughts and behaviors [7] as well as psychiatric disorders commonly associated with suicide [8]. Indeed, leading national organizations now recommend such validated screening tools for suicide risk as best practice across treatment settings [9].

Although screening questionnaires represent a key component in health care–based suicide prevention, they are not without limitations. For one, such assessments rely on patients to honestly and accurately report on their experiences and symptoms. The myriad barriers to disclosing suicidal thoughts (eg, concerns about involuntary hospitalization or other unwanted consequences and stigma) [10], as well as biases associated with retrospective recall [11], are well established. Indeed, prior studies have shown that up to three-quarters of patients who go on to die by suicide deny any suicidal thoughts in their most recent health care encounter [12]. Second, suicidal thoughts can fluctuate rapidly over short periods [13], posing the possibility that patients may not be experiencing suicidal thoughts at the time of assessment but experience an escalation in suicidal thinking soon after [10]. Third, even brief screening measures take time for patients to complete and providers to

administer and review, which may result in suboptimal completion rates and data quality [14], especially in fast-paced treatment settings. Finally, widely used screening measures have evidenced less than ideal diagnostic accuracy for predicting future suicidal behavior, with some research suggesting that even *high-risk* classifications on traditional suicide risk scales are not sufficiently accurate for clinical use [15,16]. It is recommended that validated tools be used in conjunction with clinical judgment to determine suicide risk; however, clinicians are generally quite poor at predicting who will make a suicide attempt [17]. Clearly, there is room for novel approaches to identify patients at risk for suicide in health care settings—not necessarily to replace but rather to complement traditional methods [18].

The development of automated machine learning–based models that use electronic health record (EHR) data (eg, demographic and health information) to predict patients' risk of suicidal behavior in the future is one such promising approach that has received increasing attention in the literature [19] and media [20]. There are many potential advantages of machine learning models for suicide risk prediction. For example, because such models leverage vast amounts of routinely collected clinical data, they should require little additional provider or patient burden at the point of care and do not rely exclusively on patients to accurately report on their suicidal thoughts. This also makes EHR-based models potentially useful for population-wide suicide risk screening of patients in the health care system, rather than responding only to those individuals who actively self-report suicidal thoughts or whose clinicians who identify suicide risk. Multiple research teams, including ours [21,22], have published results from studies that build and evaluate such suicide risk–prediction models. Overall, findings from this body of work are promising, with algorithms generally demonstrating high levels of accuracy (eg, classification accuracy up to 94% at Mass General Brigham, with similar model performance across 4 other large health care systems) [21,22]. Perhaps most promising is the potential to pair such scalable, low-cost models with evidence-based interventions; for example, patients stratified at the higher end of a suicide risk distribution might be prioritized for more costly and targeted clinical evaluation and monitoring and possibly receive a suicide-focused intervention [23,24]. Along these lines, recent work shows that existing suicide risk–prediction models have sufficient accuracy to be cost-effective if implemented and paired with various evidence-based interventions [25].

Despite the promise of statistical suicide risk–prediction models when used to inform clinical decision-making, with few exceptions [26-28], these efforts have yet to be widely deployed or evaluated in clinical practice. There are many complex

practical and ethical questions about how such models—and corresponding clinical decision support (CDS) tools [29] that alert providers with statistical information about patients' suicide risk and offer decision support contingent on their risk (and potentially other patient factors) at the point of care—would fare in *real-world* clinical settings that currently remain unanswered. For example, when and how frequently would suicide risk–prediction models be updated and providers notified? Would the additional workload associated with responding to automated suicide risk alerts be manageable for providers [27]? Would some machine learning (sometimes referred to as *black box*) models be interpretable or actionable [30]? Questions related to the clinical implementation of such tools also impact the patients whose data are being used to generate predictions—for example, regarding data privacy and communication model results—as well as the health care systems more broadly in which such approaches are used. A key concern, for instance, is that using automated models to identify at-risk patients who are not currently identified via clinical assessments would add more burden to health care systems with already limited resources.

Thus, before widespread clinical deployment, it is critical to partner with stakeholders (eg, frontline clinicians, patients, administrators, and payers) who can help guide such efforts. Recent work in this area has involved collecting self-report survey data from mental health professionals [30] and patients [31] on clinical and operational issues pertaining to automated suicide risk models. A recent study used self-report surveys (n=35) and interviews (n=12) to collect qualitative data from Veterans Affairs (VA) clinicians involved in the recently implemented VA program that uses predictive analytics to identify and provide outreach to veterans at high risk for suicide [32]. However, we are not aware of any published or systematic qualitative research in the form of focus groups or interviews with key stakeholders regarding the clinical implementation of suicide risk–prediction models in non-VA health care settings. Now considered essential to real-world implementation efforts, qualitative methods have the potential to both rigorously and efficiently answer key questions related to both *whether* and *how* [33] we should proceed with implementing new and automated approaches to suicide risk prediction in clinical care.

Objectives

The aim of this study is to solicit perspectives on the deployment of suicide risk–prediction models in clinical care via focus groups with providers from various departments at a large (>1000 beds, >1.5 million outpatient visits/year) urban hospital. We conducted this study to guide the ongoing development and, ultimately, clinical implementation of a CDS system that provides statistical information from suicide risk–prediction models and corresponding decision-making support to providers at the point of care. Our specific goals are to (1) better understand providers' current practices for suicide risk assessment and intervention in routine care; (2) determine providers' perspectives on the use of machine learning models and corresponding CDS tools for identifying and managing suicide risk, including barriers and facilitators; and (3) identify key factors and recommendations to consider in the development of such a CDS system.

Methods

Participants

Focus group participants were providers at the Massachusetts General Hospital (MGH) in Boston. Participants were selected via convenience sampling. The senior author (JWS) contacted leadership (eg, chiefs and clinical or training directors) of 4 MGH departments (psychiatry, internal medicine and primary care, emergency medicine, and obstetrics and gynecology) to provide a brief overview of the study and request that they identify providers in their respective departments who might be interested in participating. Leadership from these departments then either circulated (via email or face-to-face communication) the opportunity to individual providers or sent the study team names and contact information of prospective participants, whom the study team then contacted directly. The study team additionally extended several direct invitations to specific individuals within psychiatry and emergency medicine with known interest in suicide prevention or expertise in treating suicidal patients. Of the 60 total providers whom the study team contacted, 56 (93%) agreed to participate (via *implied consent* because a waiver of documentation of informed consent was obtained); of these 56 who agreed, 71% (40/56) ultimately attended a 2-hour focus group. A total of 3 focus groups comprised only primary care providers, whereas the other 7 groups included either psychiatry providers alone or psychiatry, emergency, and obstetrics and gynecology providers. Participants were paid US $250 per hour for the focus groups.

Ethics Approval

All procedures were approved under MGH Institutional Review Board 2019P001774.

Procedure

Overview

New participants were recruited until no new relevant knowledge was being obtained from the discussions (ie, data saturation achieved) for a total of 10 two-hour focus groups. All 10 groups were conducted from January through March 2020. A total of 7 groups were conducted in person (before the COVID-19 pandemic) and 3 via videoconference (during the COVID-19 pandemic; late March 2020).

Baseline Questionnaire

Upon arrival, participants were asked to complete a self-report questionnaire in REDCap (Research Electronic Data Capture) [34], a secure, web-based, electronic data capture tool. This questionnaire consisted of 30 items spanning four categories: demographic characteristics, clinical experience, EHR use (eg, number of automated EHR alerts per week), and current suicide risk assessment practices.

Focus Group Discussion

Each focus group ran for 2 hours and was led by 2 to 3 doctoral-level facilitators (KHB, KLZ, and RGF). The study coordinator (EMM) also was present for 8 of the 10 groups, and for 2 groups, 1 or 2 members of the senior author's (JWS) team observed the discussion. All focus group components after the

baseline questionnaire were audio recorded. After the baseline questionnaire, a facilitator met 1:1 with each participant for approximately 10 minutes to review their typical workflow in the EHR, with the goal of gathering information to inform the potential incorporation of a CDS system for identifying and managing suicide risk into workflows.

Next, the majority (approximately 1 hour) of each focus group was spent in an open-ended discussion assessing providers' current suicide risk assessment and intervention practices and perspectives and attitudes toward incorporating machine learning models for suicide risk prediction (and corresponding CDS tools) in clinical practice, as well as key barriers and facilitators. For this and all group discussions, facilitators used an interview guide (Multimedia Appendix 1) that was developed and refined over several months by the study team and underwent pilot testing in an initial *mock* focus group with volunteers (ie, colleagues not directly involved with the research) identified by the study team.

Finally, to elicit more specific recommendations about a potential CDS system for identifying and responding to suicide risk, providers were asked to give feedback on an initial prototype of a CDS tool that communicates statistical information (from a machine learning model using routinely collected EHR data [21,22]) about patients' suicide risk at the point of care (see the screenshot in Figure 1). Built as a webpage, the prototype consisted of a dashboard displaying demographic characteristics of a mock patient and the patient's automated *suicide risk score* (from the EHR-based model) as well as their *top risk factors* (ie, strongest predictors in the model) and absolute and relative risks of suicide attempt in the next 90 days, with links embedded to additional resources (eg, clinical suicide risk assessments). This initial prototype did not include clinical recommendations tailored to the suicide risk score. Participants were given 5 minutes to navigate the prototype on their own, after which the interview guide was used to elicit feedback in a brief (approximately 15-minute) group discussion.

Figure 1. Initial prototype of CDS system for identifying and managing suicide risk shown to participants. Name, demographics, and data shown are for a fake patient. CDS: clinical decision support.
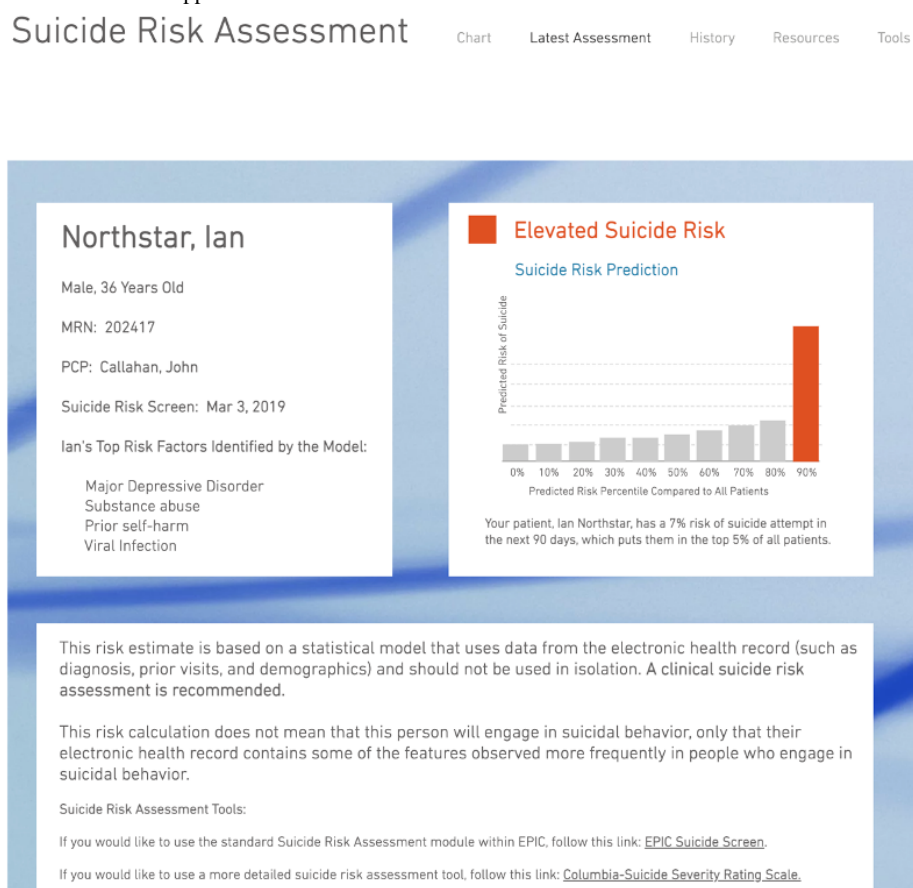
Qualitative Data Analysis

Codebook Development

Audio recordings of the group discussions were first transcribed by an external service (TranscribeMe). Then, 3 doctoral-level team members (KHB, KLZ, and RGF) independently reviewed and annotated 2 randomly selected transcripts (for a total of 6 of the 10 transcripts) to inductively derive major and recurrent

themes related to key guiding interview questions (Multimedia Appendix 1). Themes were then organized into an initial draft of a codebook, and a comprehensive training manual with coding guidelines was provided to study the staff responsible for coding. Five team members (KHB, RGF, EMM, HL, and DK) then each used the draft codebook to practice coding another transcript using NVivo (version 12) [35] with the goal of identifying new codes to add or irrelevant or redundant codes to delete.

The results of this practice coding and potential new or redundant codes were reviewed, as well as coding issues troubleshooted, in a series of team meetings. Whereas some codes were removed, others were merged or renamed for clarity. The codebook was then finalized to contain 140 unique themes (each with example quotes), organized into the following eight subcategories: (1) current suicide risk assessment practices, (2) current suicide risk intervention practices, (3) attitudes about current assessment and interventions, (4) general reactions and attitudes toward using machine learning models for suicide risk prediction in clinical practice, (5) factors to consider when developing or implementing such models and corresponding CDS tools for suicide risk, (6) barriers and concerns to using such systems, (7) recommendations about system content or format, and (8) recommendations about placement of systems within the EHR.

Coding Process

Each of the 10 transcripts was then independently reviewed and coded by 2 of the 3 study coders (EMM, HL, and DK) in NVivo. All coded text was reviewed, and discrepancies were resolved in consensus meetings among the 3 coders and at least one doctoral-level team member who facilitated the focus groups. To account for the fact that some participants were more talkative than others and repeated the same theme across multiple distinct consecutive statements, the same theme was only coded once when the same speaker made the same point multiple times within a 2-minute time span. The final coded versions of each transcript, with consensus achieved on all discrepancies, were used to generate the findings (ie, recurrent, frequently coded individual themes or clusters of individual themes) presented in the *Results* section.

Results

Sample Characteristics and Questionnaire Data

The mean age of the participants was 43.1 (SD 12.1; median 40.0) years. Of the 40 participants, 15 (38%) were aged between 25 and 34 years, 7 (18%) between 35 and 44 years, 9 (23%) between 45 and 54 years, 7 (18%) between 55 and 64 years, and 2 (5.0%) between 65 and 74 years. The majority were women (27/40, 68%), with 33% (13/40) identifying as men. Participants were predominantly White (35/40, 88%), followed by Asian (2/40, 5%); of the 40 participants, 1 (3%) was Black or African American, 1 (3%) identified as belonging to >1 race, 1 (3%) preferred not to answer, and 1 (3%) was Hispanic or Latino. Of the 40 providers, 24 (60%) were from psychiatry, 10 (25%) were from internal medicine and primary care, 5 (13%) were from emergency medicine, and 1 (3%) was from obstetrics and gynecology. In addition, of the 40 participants, 28 (70%) were physicians, 10 (25%) were psychologists, 1 (3%) was a social worker, and 1 (3%) was a nurse practitioner. Less than one-third (11/40, 28%) identified as residents or trainees. Providers reported an average of 15.1 (SD 10.9) years of experience treating patients (median 10.5 years); 44% (17/39) of participants reported between 0 and 9.9 years treating patients, 15% (6/39) between 10 and 19.9 years, 26% (10/39) between 20 and 29.9 years, and 15% (6/39) reported ≥30 years. All physicians and psychologists reported working in outpatient settings and 63% (25/40) also in inpatient, emergency, or urgent care settings. All internal medicine and primary care physicians (PCPs) reported working in outpatient settings and just less than one-third (12/40, 30%) also in inpatient or urgent care settings. More than half (22/40, 55%) of the participants reported spending ≥33 hours per week on direct patient care.

More than two-fifths (17/40, 43%) of the participants reported using an EHR for ≥10 years. Just more than one-third (14/40, 35%) of the participants received >20 automated alerts each week, and more than half (21/40, 53%) of the participants reported receiving *too many* automated EHR alerts. It was most common for providers to report feeling *somewhat* burned out (17/40, 43%) by their clinical work, and 30% (12/40) felt *a little bit* burned out. Similarly, providers were most likely to report feeling *somewhat* (21/40, 53%) or *a little bit* (7/40, 18%) burned out by their use of the EHR. Just more than half (21/40, 53%) of the participants reported documenting a patient's risk for suicide at every visit, followed by 38% (15/40) at less than half of patient visits. The majority (27/40, 68%) of the participants reported commonly using at least one structured or semistructured tool to assess suicide risk: most often, the Patient Health Questionnaire-9 [8] or the structured suicide risk assessment tool embedded in the EHR (both endorsed by 22/40, 55% of participants), followed by the Columbia-Suicide Severity Rating Scale [7] (12/40, 30% of participants).

Themes From Group Discussion

Both the count of coded themes (across the 10 focus groups) and proportion of groups in which themes were coded are provided in Table 1. The top 5 coded themes in each category are displayed in Figure S1 (Multimedia Appendix 2). Example quotes are presented in Table S1 in Multimedia Appendix 3.

Table 1. Counts and percentages of frequently coded themes.

Theme	Times theme was coded	Groups with coded theme, n (%)
Current suicide risk assessment and intervention practices		
Other risk assessment practices (eg, review EHR[a] and obtain collateral)	56	10 (100)
Use unstructured clinical interviewing	42	10 (100)
Use structured or semistructured tools	41	9 (90)
Consult with colleague, supervisor, or external service	26	9 (90)
Refer for emergency evaluation or inpatient hospitalization	22	9 (90)
Assessing or predicting suicide risk is challenging or frustrating	17	7 (70)
Do not use structured or semistructured tools	17	7 (70)
Structured or semistructured tools are unhelpful, or clinical interviewing best	13	8 (80)
Access problem in mental health treatment	13	7 (70)
Time constraints (associated with thorough suicide risk assessment)	13	6 (60)
Risk and liability concerns for providers treating suicidal patients	12	6 (60)
Refer to on-site mental health support	12	4 (40)
Connect with patient's current psychiatry provider	11	6 (60)
Low threshold for consulting psychiatry (includes sending to ED[b] for evaluation)	10	5 (50)
Behavioral health team not accessible for consults or supports	9	3 (30)
Lack of comfort with suicide risk assessment or current practices	8	4 (40)
Value of on-site (or consulting) behavioral health presence	8	2 (20)
Structured or semistructured (or mandated) tools are helpful	5	4 (40)
Develop safety plan or other brief suicide-focused intervention	2	2 (20)
General attitudes about automated suicide risk–prediction models		
General interest or promise, or would trust once implemented	47	9 (90)
Interest in using tool at point of care or as a BPA[c]	38	10 (100)
General skepticism, sounds anxiety-provoking, or would not trust	33	9 (90)
Must outperform clinical judgment or show accuracy before clinical use	33	8 (80)
Promise in primary care	24	8 (80)
Promise for identifying high-risk patients who might otherwise be missed	15	7 (70)
Promise for population-level risk stratification (and resource allocation)	12	7 (70)
Promise in ED (or psychiatric ED)	9	7 (70)
Promise for new evaluations and certain types of patients	9	6 (60)
Little or no promise in ED (or psychiatric ED)	5	4 (40)
Barriers and concerns		
Liability	39	10 (100)
Low data quality in EHR	19	6 (60)
Alert fatigue or desensitization to suicide risk alerts	18	7 (70)
Increase in rates of patients needing emergency evaluations or inpatient beds	18	6 (60)
Increase access problem in psychiatry or contribute to overall system burden	16	5 (50)
Other harmful effects for patients (eg, stigma and provider-patient alliance)	15	8 (80)
Utility depends on interventions that would be triggered	14	5 (50)
Potential for alert to come during a visit unrelated to mental health	14	5 (50)
Time constraints (associated with using additional CDS[d] tool)	10	5 (50)

Theme	Times theme was coded	Groups with coded theme, n (%)
How to respond to risk communicated by tool outside of face-to-face visits	9	4 (40)
Facilitators and specific recommendations		
Interest in viewing patients' predictors or features contributing to risk score	49	10 (100)
Must have good user interface and user experience	37	9 (90)
Want to see changes in risk scores over time	34	9 (90)
Need for standardized workflows for responding and documentation	29	8 (80)
Information should be available to all patients' providers	18	7 (70)
Importance of provider training (including instruction if tool is mandatory)	17	6 (60)
Should be pushed to, not pulled by, provider	14	7 (70)
Want more information on how algorithm works or test characteristics	12	8 (80)
Prompt further assessment with structured or semistructured tools or specific questions	11	5 (50)
Use tool in combination with clinical judgment	11	5 (50)
Should distinguish between chronic and time-varying predictors or features	10	6 (60)
Do not recommend interventions for specific risk scores or features	9	3 (30)
Information in tool should not be available to others with EHR access	7	4 (40)
Give recommendations of interventions for specific risk scores or features	7	4 (40)
Should be pulled by, not pushed to, provider	3	2 (20)
Other factors to consider		
Patients should be able to see information in the tool	21	6 (60)
Timing of when information is given to provider	20	7 (70)
Importance of considering whether patients will see this information	18	7 (70)
Risk-prediction window	16	7 (70)
Patients should not have access to the information in the tool	14	6 (60)
Variability in interventions, thresholds, resources, or EHR use across settings	10	5 (50)

[a]EHR: electronic health record.

[b]ED: emergency department.

[c]BPA: best practice advisory.

[d]CDS: clinical decision support.

Current Suicide Risk Assessment and Interventions

All types of providers most often described using unstructured clinical interviewing (not necessarily in isolation from other assessment methods) to determine suicide risk. Although it was also very common for providers to report using structured or semistructured tools, there were also providers in most groups who stated that they do not use structured or semistructured risk assessment tools. Other common risk assessment practices included reviewing the EHR for relevant historical information, considering known suicide risk factors, obtaining information from collaterals, assessing access to lethal means, and using clinical observation (eg, mental status examination).

Regarding current interventions, providers most commonly described consulting with a colleague or supervisor. It was also common for providers to indicate referring high-risk patients for emergency evaluation or inpatient hospitalization, with a subset having a low threshold for doing so. Providers often acknowledged the access problem in mental health treatment.

For PCPs, referring patients to an on-site mental health professional for evaluation or support was often reported, as well as connecting with a patient's current mental health provider (if they have one). Notably, developing a safety plan or other brief, suicide-focused interventions with empirical support [24] was only mentioned in 20% (2/10) of the groups.

Providers expressed a variety of attitudes about their current practices for assessing and treating suicide risk. Predominantly, participants referred to assessing suicide risk or predicting whether a patient will make a suicide attempt as challenging or frustrating; notably, this was more common among providers in psychiatry and emergency medicine than in primary care. Despite how often structured or semistructured tools are used, it was more common for providers to describe these tools as unhelpful than helpful. Expressing concerns about the risk and liability associated with managing patients at risk for suicide was also fairly common. Some providers also noted the time constraints during visits that can make thoroughly assessing suicide risk difficult, and others expressed an overall lack of

comfort with suicide risk assessment and intervention practices. PCPs emphasized both the value of on-site mental health professionals for consultations or support *and* that they are not as available or accessible as would be ideal.

General Attitudes About Automated Suicide Risk–Prediction Models

The predominant attitude toward the use of automated suicide risk–prediction models in clinical practice was positive, with providers generally highlighting their overall interest in this approach and potential to trust model predictions once implemented in routine clinical care. However, many also emphasized the importance of demonstrating that such a tool outperforms or supplements clinical judgment and is accurate before clinical use. It was less common, but still mentioned in nearly all of the focus groups, for providers to also express varying degrees of skepticism, such as that they would either distrust such a tool or find it anxiety-provoking.

Regarding specific potential use cases, providers (especially PCPs) noted the potential promise to inform clinical decision-making and treatment planning at the point of care (eg, as a best practice advisory) and, to a lesser extent, for purposes of population-level risk stratification and resource allocation. The latter was indicated more often by providers in psychiatry and emergency medicine. Providers were especially interested in the potential for the tool to identify high-risk patients who could be missed via traditional assessments. Providers often noted that the value of suicide risk–prediction models would vary by treatment setting, with primary care as the most optimal, followed by the ED. Less commonly, others explicitly emphasized that such a tool would not be useful in the ED. It was also noted in a few groups that such a tool may have more value during new evaluations than during follow-up visits and with certain types of patients.

Barriers and Concerns

The single most common barrier to the potential use of automated suicide risk–prediction models in routine care was the implication for liability. For example, many were concerned about being held legally responsible if they decided not to hospitalize a patient who was categorized as high risk and then went on to attempt or die by suicide. It was also fairly common to express concerns about low data quality in the EHR affecting the quality of model output (ie, *garbage in, garbage out*). Another frequent concern was the potential for alerts generated by automated suicide risk–prediction models to increase alert fatigue and, specific to this application, that providers may become desensitized to suicide risk alerts over time.

Providers were also concerned about the potential for such a tool to lead to increased rates of hospitalizing patients in health care systems already facing ED overcrowding and shortages of inpatient beds. The possibility for such a tool to further increase the (currently unmet) demand for outpatient mental health services was often raised, and participants noted that the system must offer additional resources if this tool were implemented. Indeed, some providers emphasized that the usefulness of the tool would depend on what *next steps* (eg, interventions or referrals) are available. Whether there might be other associated

harmful effects for patients (eg, stigma) was a concern frequently voiced by psychiatry and ED providers. It was especially common for PCPs (less so other providers) to question the benefits (and actionability) of learning that a patient is at high risk for suicide during an encounter unrelated to mental health (eg, ED visit for a medical reason). Other less frequently noted barriers included time constraints and issues associated with responding to suicide risk alerts if received outside of visits.

Facilitators and Specific Recommendations

Providers strongly believed that such a tool must have a good user interface and user experience to be helpful. Providers also overwhelmingly described an interest in being able to view the specific predictors (eg, diagnoses, demographic characteristics, and treatment attendance) that contribute to an individual patient being identified as high risk for suicide via the model, with some who encouraged distinguishing between chronic and time-varying predictors (or *risk factors*) within the tool. Providers were also extremely interested in being able to view *changes* in patients' suicide risk scores over time. The need to establish clear, standardized workflows for both responding to suicide risk predictions generated by the tool and documenting interactions with the tool was also a common theme. Similarly, providers emphasized the importance of receiving systematic training on how to use the tool before it is rolled out clinically, including instruction on whether or not its use is mandatory.

It was more common for providers to state that information about suicide risk generated by the tool should be available to all treaters (across disciplines and departments) for a given patient and less common to indicate that model predictions should not be viewed by certain types of providers (eg, physical therapists and cardiologists) and other medical or administrative staff (eg, medical assistants and receptionists). It was generally preferred that suicide risk scores (or alerts) be *pushed to*, rather than *pulled by*, the provider. Some requested having information available within the tool on how the algorithm works and key test characteristics (eg, sensitivity and specificity). It was fairly common for providers to suggest that the tool offer specific follow-up questions to ask, or standardized scales to use, for additional (clinical) suicide risk assessment. However, it was slightly more common to express the view that the tool should not recommend specific interventions based on risk score or primary predictors as opposed to the view that the tool *should* recommend specific interventions. Non-PCPs (eg, psychiatry or ED) often emphasized that the tool be used in combination with (not as a substitute for) clinical judgment.

Other Factors to Consider

Other factors that participants recommended be considered included the timing of when statistical information about suicide risk is delivered or made available to providers (eg, before, at the start of, or during visits, particularly given the possibility of no-shows) as well as the temporal window used for model predictions (eg, risk over the next week, month, or year). Providers generally believed it would be vital to decide early on whether patients will have access to their suicide risk scores, as this could have implications for what information is included in the tool and wording of the text. It was more common for providers to express the view that patients should (vs should

not) have access to model predictions. Providers also noted that there is wide variability across treatment settings in interventions available, conventions of EHR use, and relevant risk thresholds (eg, to warrant categorizing a patient as *high risk* for suicide) and that taking this into account would be important to consider when developing a tool intended for use across the health care system.

Discussion

Principal Findings

Interest in machine learning models for suicide risk prediction has proliferated in recent years. The potential to use such models and corresponding CDS systems in routine care is compelling; however, before incorporation into practice, it is critical to partner with key stakeholders who can offer the perspectives and feedback necessary for successful clinical implementation. This focus group study with hospital-based providers revealed several key areas of findings that have direct implications for ongoing and future development and deployment efforts.

First, providers were not satisfied with currently available suicide risk assessment methods. For example, despite the fact that most providers endorsed using structured or semistructured measures of suicide risk, existing tools were more often described as not particularly helpful rather than helpful. However, it is important to consider that this finding in the context of providers also tended to report feeling burned out by their clinical work, posing the possibility that negative attitudes toward current assessments are driven in part by burnout about their job in general (rather than only a specific dislike for current tools). Although unstructured clinical interviewing was generally viewed as the best available method for determining suicide risk, recent research suggests that providers tend to be quite poor at predicting the risk of suicidal behavior [17]. Indeed, many providers referred to how challenging and frustrating it can be to determine a patient's risk for suicide. Taken together, these findings underscore the clear opportunity to improve upon existing, traditional methods of suicide risk assessment.

Encouragingly, reactions to incorporating machine learning models for suicide risk prediction in clinical practice were positive overall. These approaches appeared especially promising to providers if used to identify and notify providers about patients at the high end of the suicide risk distribution either at the point of care (eg, to inform clinical decision-making) or for purposes of population-level resource allocation, meaning that those identified as higher-risk would be prioritized to receive suicide-specific interventions or more costly or difficult-to-access treatment and follow-up. Our recent work suggests that suicide risk–prediction models may have especially good accuracy when identifying patients at both the very high *and* very low ends of the risk distribution [36]. Given this, it may be worthwhile to frame and leverage these models such that they are also clinically useful for providers seeing patients who are very *unlikely* to go on to make a suicide attempt (and thus may not require further intervention).

Some providers also expressed skepticism, a lack of trust, or anxiety about the potential use of automated suicide risk

assessment models in clinical practice, with liability as the single most discussed concern about using these methods in practice. These findings suggest the importance of partnering with individuals and organizations who specialize in medical risk management, as well as with other relevant legal and payer stakeholders, as these tools are brought closer to clinical use. Overarching concerns about liability may be somewhat mitigated by emphasizing both the importance of using model predictions in conjunction with clinical judgment and that high-risk scores should be interpreted in the context of all other available information to make treatment or discharge decisions. Broadly, developing standards of care accompanied by systematic psychoeducation, training, and protocols to accompany such CDS tools for the providers who will ultimately use them may be critical to address concerns about liability and foster buy-in and promote confidence in their use.

Another concern was that alerts from such a CDS tool that incorporates suicide risk–prediction models would be disruptive to providers' workflows and result in alert fatigue. To maximize the clinical utility and cost-effectiveness of machine learning–based suicide risk models, empirically derived risk thresholds that balance the relative value of avoiding false-negative and false-positive errors [25,27] must be established. Indeed, recent work in the Kaiser Permanente health care system suggests that the degree to which suicide risk models add to clinical workloads depends heavily on the risk threshold selected, along with the approach for responding to these alerts [27]. Such thresholds may also vary by setting; for example, in psychiatry, a higher threshold may be preferred to reduce the number of false positives.

An overarching concern was that EHR-based suicide risk models and corresponding CDS tools would result in an increased burden on already overburdened health care systems. This could happen via either increased rates of sending patients to the ED for further evaluation or inpatient hospitalizations (possibly in part driven by provider anxiety and liability concerns) or more demand for outpatient specialty services. This is consistent with the fact that providers noted already having a low threshold to section patients and problems accessing mental health treatment. Moreover, as some providers stated that the tool's usefulness would depend on what interventions are available to them, health systems may benefit from determining the approximate number of *newly identified* (via the model) patients who will require interventions or referrals (eg, through simulation studies) and whether current resources can handle potential increases in demand. Developing new resources (eg, urgent care clinics and suicide-focused treatment options) or partnering with organizations outside the hospital system (eg, crisis services and referral programs) may be needed before clinical implementation.

In addition to practical concerns, providers also overwhelmingly voiced interest in viewing the individual predictors or risk factors that contribute to a patient's elevated risk score. Providers suggested that this information may also guide next steps for intervention. This suggestion is in line with recent research [30] showing that providers may be less likely to use the information from suicide risk–prediction models if the relevant clinical features are hidden *or* they do not view the

provided features as intuitively connected to suicide risk. Thus, there appear to be both pros and cons for sharing information on the specific features in these models. Displaying these features could also lead to misinterpretations about causality (eg, the assumption that if a provider targets a certain feature, suicide risk will decrease), which would not accurately reflect the complex, interactive nature of features in machine learning models.

Providers also raised various complicated yet important-to-consider issues such as the timing of when risk alerts would be delivered to providers (and corresponding concerns about, eg, receiving a *high-risk alert* outside of a visit or when a patient does not attend a scheduled appointment) and risk-prediction windows. Although there is great interest in improving short-term suicide risk prediction (eg, over hours, days, or weeks) [37], the degree to which EHR-based models accurately predict short-term risk and are sensitive to change may be limited. In the future, in addition to the clinical questionnaire data that some models incorporate [38], combining EHR-based data with fine-grained *real-time* data (eg, from smartphones or wearable devices [39]) that may capture more acute increases in risk may help meet this need. In the meantime, balancing clinical utility with the accuracy of different risk-prediction windows is critical.

Providers had mixed views on who should have access to the results from automated suicide risk–prediction models. Some felt it was important for all treaters of a given patient to access these results, whereas others thought there may be certain providers for whom such information would be irrelevant or unnecessary and who, for example, may lack training and experience with approaching the issue of suicide sensitively. Providers were also mixed as to whether patients should be able to receive this information. Given the increasing momentum of the OpenNotes movement [40], including within the mental health domain [41], it seems likely that over the long term, we can expect more EHR information to be available to service users. Thus, partnering with patient stakeholders [31] to develop protocols for transparent, sensitive communications about suicide risk, including key contextual information (eg, how to interpret being in a *high-risk* category) and language, may also be strategic.

Limitations

There are several notable limitations of this study. First, our findings reflect only the views of providers from a single academic medical center in the Northeastern United States and may not generalize outside of this clinical context. Second, the vast majority of participants were White, and thus, the findings of this study may not generalize to providers who identify as minorities. Third, convenience sampling was used for recruitment, which may also limit generalizability; for instance,

it is possible that participants were more enthusiastic about suicide risk–prediction models because some were selected for their known interest or expertise in suicide prevention. Fourth, we did not systematically assess providers' prior experience with suicide risk–prediction models, which may influence attitudes toward these tools. We were also unable to systematically disaggregate theme frequency by provider type or clinical setting, which may obscure important differences in perspectives on using suicide risk–prediction models in practice.

Future Directions

There are a few key priorities for future investigation. First, as the development of CDS tools for suicide risk prediction is ongoing, qualitative data must be collected both on more fully developed prototypes and during or after initial clinical implementation of such tools. Our emphasis in this study's series of focus groups was on general attitudes and perspectives toward such approaches, whereas future work will focus on tool development, refinement, and implementation. Second, given the potential scalability of suicide risk–prediction models, future work must include the full range of providers working in a broader range of urban and more rural hospitals nationwide, as well as community health centers and other nontraditional health care settings (both psychiatric and other medical settings, particularly given the enthusiasm from providers here about potential use in primary care). Relatedly, as noted earlier, patient stakeholders must be included in the process of developing and implementing these new tools. More work is also needed to determine how health care systems and individual providers may best leverage and combine information from machine learning models and traditional clinical assessments [18,36]. Finally, given recent work suggesting the potential for suicide risk–prediction models to exacerbate existing racial or ethnic disparities [42], upcoming work should explicitly probe providers' (and other stakeholders') perspectives on how to address this critical issue before clinical implementation.

Conclusions

Deploying virtually any new tool in practice has challenges, but given the complex, sensitive, and unfortunately still stigmatized nature of suicide, it is especially vital that researchers working in this area involve the full range of stakeholders in each stage of the implementation process. Overall, providers in the current qualitative study were dissatisfied with current suicide risk assessment methods and were open to the use of a machine learning–based suicide risk identification and management system to inform their clinical decision-making. This work highlights several key potential barriers and facilitators to be addressed in ongoing and future efforts to develop and implement such models and corresponding CDS tools in routine care and thus potentially better prevent patient suicides within health care systems.

Conflicts of Interest

Multimedia Appendix 1
Interview guide (abridged): prompts to guide group discussion.
[DOCX File , 18 KB - formative_v6i3e30946_app1.docx]

Multimedia Appendix 2
Top 5 frequently coded themes in each category.
[DOCX File , 251 KB - formative_v6i3e30946_app2.docx]

Multimedia Appendix 3
Example quotes for frequently coded themes.
[DOCX File , 17 KB - formative_v6i3e30946_app3.docx]

References

1. Preventing suicide: a global imperative. World Health Organization. 2014. URL: https://apps.who.int/iris/bitstream/handle/10665/131056/9789241564779_eng.pdf [accessed 2022-03-02]
2. Morbidity and Mortality Weekly Report: changes in suicide rates — United States, 2018–2019. Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/mmwr/volumes/70/wr/pdfs/mm7008a1-H.pdf [accessed 2022-03-02]
3. Web-based Injury Statistics Query and Reporting System (WISQARS). Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/injury/wisqars/index.html [accessed 2022-03-02]
4. Project 2025. American Foundation for Suicide Prevention. 2019. URL: https://project2025.afsp.org [accessed 2022-03-02]
5. Luoma JB, Martin CE, Pearson JL. Contact with mental health and primary care providers before suicide: a review of the evidence. Am J Psychiatry 2002 Jun;159(6):909-916 [FREE Full text] [doi: 10.1176/appi.ajp.159.6.909] [Medline: 12042175]
6. Ahmedani BK, Simon GE, Stewart C, Beck A, Waitzfelder BE, Rossom R, et al. Health care contacts in the year before suicide death. J Gen Intern Med 2014 Jun 25;29(6):870-877 [FREE Full text] [doi: 10.1007/s11606-014-2767-3] [Medline: 24567199]
7. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry 2011 Dec;168(12):1266-1277 [FREE Full text] [doi: 10.1176/appi.ajp.2011.10111704] [Medline: 22193671]
8. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
9. Recommended standard care for people with suicide risk: making health care suicide safe. National Action Alliance for Suicide Prevention. 2018. URL: https://theactionalliance.org/sites/default/files/action_alliance_recommended_standard_care_final.pdf [accessed 2022-03-02]
10. Richards JE, Whiteside U, Ludman EJ, Pabiniak C, Kirlin B, Hidalgo R, et al. Understanding why patients may not report suicidal ideation at a health care visit prior to a suicide attempt: a qualitative study. Psychiatr Serv 2019 Jan 01;70(1):40-45. [doi: 10.1176/appi.ps.201800342] [Medline: 30453860]
11. Fredrickson B, Kahneman D. Duration neglect in retrospective evaluations of affective episodes. J Personal Soc Psychol 1993;65(1):45-55 [FREE Full text] [doi: 10.1037/0022-3514.65.1.45]
12. Berman AL. Risk factors proximate to suicide and suicide risk assessment in the context of denied suicide ideation. Suicide Life Threat Behav 2018 Jun 21;48(3):340-352. [doi: 10.1111/sltb.12351] [Medline: 28429385]
13. Kleiman EM, Turner BJ, Fedor S, Beale EE, Huffman JC, Nock MK. Examination of real-time fluctuations in suicidal ideation and its risk factors: results from two ecological momentary assessment studies. J Abnorm Psychol 2017 Aug;126(6):726-738. [doi: 10.1037/abn0000273] [Medline: 28481571]
14. Vannoy SD, Robins LS. Suicide-related discussions with depressed primary care patients in the USA: gender and quality gaps. A mixed methods analysis. BMJ Open 2011 Jan 01;1(2):e000198 [FREE Full text] [doi: 10.1136/bmjopen-2011-000198] [Medline: 22021884]
15. Carter G, Milner A, McGill K, Pirkis J, Kapur N, Spittal MJ. Predicting suicidal behaviours using clinical instruments: systematic review and meta-analysis of positive predictive values for risk scales. Br J Psychiatry 2018 Jan 02;210(6):387-395. [doi: 10.1192/bjp.bp.116.182717]

16. Quinlivan L, Cooper J, Davies L, Hawton K, Gunnell D, Kapur N. Which are the most useful scales for predicting repeat self-harm? A systematic review evaluating risk scales using measures of diagnostic accuracy. BMJ Open 2016 Feb 12;6(2):e009297 [FREE Full text] [doi: 10.1136/bmjopen-2015-009297] [Medline: 26873046]

17. Woodford R, Spittal MJ, Milner A, McGill K, Kapur N, Pirkis J, et al. Accuracy of clinician predictions of future self-harm: a systematic review and meta-analysis of predictive studies. Suicide Life Threat Behav 2019 Feb 03;49(1):23-40. [doi: 10.1111/sltb.12395] [Medline: 28972271]

18. Simon GE, Matarazzo BB, Walsh CG, Smoller JW, Boudreaux ED, Yarborough BJ, et al. Reconciling statistical and clinicians' predictions of suicide risk. Psychiatr Serv 2021 May 01;72(5):555-562. [doi: 10.1176/appi.ps.202000214] [Medline: 33691491]

19. Belsher BE, Smolenski DJ, Pruitt LD, Bush NE, Beech EH, Workman DE, et al. Prediction models for suicide attempts and deaths. JAMA Psychiatry 2019 Jun 01;76(6):642. [doi: 10.1001/jamapsychiatry.2019.0174]

20. Carey B. Can an algorithm prevent suicide? The New York Times. 2020. URL: https://www.nytimes.com/2020/11/23/health/artificial-intelligence-veterans-suicide.html [accessed 2022-03-02]

21. Barak-Corren Y, Castro VM, Javitt S, Hoffnagle AG, Dai Y, Perlis RH, et al. Predicting suicidal behavior from longitudinal electronic health records. Am J Psychiatry 2017 Feb 01;174(2):154-162. [doi: 10.1176/appi.ajp.2016.16010077] [Medline: 27609239]

22. Barak-Corren Y, Castro VM, Nock MK, Mandl KD, Madsen EM, Seiger A, et al. Validation of an electronic health record-based suicide risk prediction modeling approach across multiple health care systems. JAMA Netw Open 2020 Mar 02;3(3):e201262 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.1262] [Medline: 32211868]

23. D'Anci KE, Uhl S, Giradi G, Martin C. Treatments for the prevention and management of suicide. Ann Intern Med 2019 Aug 27;171(5):334. [doi: 10.7326/m19-0869]

24. Doupnik SK, Rudd B, Schmutte T, Worsley D, Bowden CF, McCarthy E, et al. Association of suicide prevention interventions with subsequent suicide attempts, linkage to follow-up care, and depression symptoms for acute care settings. JAMA Psychiatry 2020 Oct 01;77(10):1021. [doi: 10.1001/jamapsychiatry.2020.1586]

25. Ross EL, Zuromski KL, Reis BY, Nock MK, Kessler RC, Smoller JW. Accuracy requirements for cost-effective suicide risk prediction among primary care patients in the US. JAMA Psychiatry 2021 Jun 01;78(6):642. [doi: 10.1001/jamapsychiatry.2021.0089]

26. Kessler RC, Hwang I, Hoffmire CA, McCarthy JF, Petukhova MV, Rosellini AJ, et al. Developing a practical suicide risk prediction model for targeting high‐risk patients in the Veterans health Administration. Int J Methods Psychiatr Res 2017 Jul 04;26(3):e1575. [doi: 10.1002/mpr.1575]

27. Kline-Simon AH, Sterling S, Young-Wolff K, Simon G, Lu Y, Does M, et al. Estimates of workload associated with suicide risk alerts after implementation of risk-prediction model. JAMA Netw Open 2020 Oct 21;3(10):e2021189. [doi: 10.1001/jamanetworkopen.2020.21189]

28. Walsh CG, Johnson KB, Ripperger M, Sperry S, Harris J, Clark N, et al. Prospective validation of an electronic health record-based, real-time suicide risk model. JAMA Netw Open 2021 Mar 01;4(3):e211428 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.1428] [Medline: 33710291]

29. Moja L, Kwag KH, Lytras T, Bertizzolo L, Brandt L, Pecoraro V, et al. Effectiveness of computerized decision support systems linked to electronic health records: a systematic review and meta-analysis. Am J Public Health 2014 Dec;104(12):12-22. [doi: 10.2105/ajph.2014.302164]

30. Brown LA, Benhamou K, May AM, Mu W, Berk R. Machine learning algorithms in suicide prevention. J Clin Psychiatry 2020 Apr 21;81(3):19m12970. [doi: 10.4088/jcp.19m12970]

31. Yarborough BJ, Stumbo SP. Patient perspectives on acceptability of, and implementation preferences for, use of electronic health records and machine learning to identify suicide risk. Gen Hosp Psychiatry 2021 May;70:31-37. [doi: 10.1016/j.genhosppsych.2021.02.008] [Medline: 33711562]

32. Piccirillo M, Pruitt L, Reger M. Evaluating clinician attitudes after local implementation of the veterans affairs predictive analytic model for suicide prevention. J Psychiatr Pract 2022 Jan 06;28(1):14-23. [doi: 10.1097/PRA.0000000000000595] [Medline: 34989341]

33. Hamilton AB, Finley EP. Qualitative methods in implementation research: an introduction. Psychiatry Res 2019 Oct;280:112516 [FREE Full text] [doi: 10.1016/j.psychres.2019.112516] [Medline: 31437661]

34. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]

35. Computer software (Version 12). QSR International Pty Ltd. 2018. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2022-03-02]

36. Nock MK, Millner AJ, Ross EL, Kennedy CJ, Al-Suwaidi M, Barak-Corren Y, et al. Prediction of suicide attempts using clinician assessment, patient self-report, and electronic health records. JAMA Netw Open 2022 Jan 04;5(1):e2144373 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.44373] [Medline: 35084483]

37. Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. Psychol Bull 2017 Feb;143(2):187-232. [doi: 10.1037/bul0000084] [Medline: 27841450]

38. Simon GE, Johnson E, Lawrence JM, Rossom RC, Ahmedani B, Lynch FL, et al. Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records. Am J Psychiatry 2018 Oct 01;175(10):951-960. [doi: 10.1176/appi.ajp.2018.17101167] [Medline: 29792051]

39. Torous J, Larsen ME, Depp C, Cosco TD, Barnett I, Nock MK, et al. Smartphones, sensors, and machine learning to advance real-time prediction and interventions for suicide prevention: a review of current progress and next steps. Curr Psychiatry Rep 2018 Jun 28;20(7):51. [doi: 10.1007/s11920-018-0914-y] [Medline: 29956120]

40. Walker J, Leveille S, Bell S, Chimowitz H, Dong Z, Elmore JG, et al. Opennotes after 7 years: patient experiences with ongoing access to their clinicians' outpatient visit notes. J Med Internet Res 2019 May 06;21(5):e13876 [FREE Full text] [doi: 10.2196/13876] [Medline: 31066717]

41. Blease CR, O'Neill S, Walker J, Hägglund M, Torous J. Sharing notes with mental health patients: balancing risks with respect. Lancet Psychiatry 2020 Nov;7(11):924-925. [doi: 10.1016/s2215-0366(20)30032-8]

42. Coley RY, Johnson E, Simon GE, Cruz M, Shortreed SM. Racial/ethnic disparities in the performance of prediction models for death by suicide after mental health visits. JAMA Psychiatry 2021 Jul 01;78(7):726-734. [doi: 10.1001/jamapsychiatry.2021.0493] [Medline: 33909019]

Abbreviations

CDS: clinical decision support
ED: emergency department
EHR: electronic health record
MGH: Massachusetts General Hospital
PCP: primary care physician
PI: principal investigator
VA: Veterans Affairs

Original Paper

Evaluation of a Web-Based Dietary Assessment Tool (myfood24) in Norwegian Women and Men Aged 60-74 Years: Usability Study

Laila Arnesdatter Hopstock[1], BSc, CRNA, MSc, PhD; Anine Christine Medin[2], BSc, MSc, PhD; Guri Skeie[1], BSc, MSc, PhD; André Henriksen[3], BSc, MSc, PhD; Marie Wasmuth Lundblad[1], BSc, MPH, PhD

[1]Department of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway

[2]Department of Nutrition and Public Health, University of Agder, Kristiansand, Norway

[3]Department of Computer Science, UiT The Arctic University of Norway, Tromsø, Norway

Corresponding Author:
Laila Arnesdatter Hopstock, BSc, CRNA, MSc, PhD
Department of Community Medicine
UiT The Arctic University of Norway
Hansine Hansens vei
Tromsø, 9037
Norway
Phone: 47 77644000
Email: laila.hopstock@uit.no

Abstract

Background: A healthy diet throughout the life course improves health and reduces the risk of disease. There is a need for new knowledge of the relation between diet and health, but existing methods to collect information on food and nutrient intake have their limitations. Evaluations of new tools to assess dietary intake are needed, especially in old people, where the introduction of new technology might impose challenges.

Objective: We aimed to examine the usability of a new web-based dietary assessment tool in older adult women and men.

Methods: A total of 60 women and men (participation 83%, 57% women) aged 60-74 years recruited by convenience and snowball sampling completed a 24-hour web-based dietary recall using the newly developed Norwegian version of Measure Your Food On One Day (myfood24). Total energy and nutrient intakes were calculated in myfood24, primarily on the basis of the Norwegian Food Composition Table. No guidance or support was provided to complete the recall. Usability was assessed using the system usability scale (SUS), where an SUS score of ≥68 was considered satisfactory. We examined the responses to single SUS items and the mean (SD) SUS score in groups stratified by sex, age, educational level, and device used to complete the recall (smartphone, tablet device, or computer).

Results: The mean total energy intake was 5815 (SD 3093) kJ. A total of 14% of participants had an energy intake of <2100 kJ (ie, 500 kilocalories) and none had an intake of >16,800 kJ (ie, 4000 kilocalories). Mean energy proportions from carbohydrates, fat, protein, alcohol, and fiber was within the national recommendations. The mean SUS score was 55.5 (SD 18.6), and 27% of participants had SUS scores above the satisfactory product cut-off. Higher SUS scores were associated with younger age and lower education, but not with the type of device used.

Conclusions: We found the overall usability of a new web-based dietary assessment tool to be less than satisfactory in accordance with standard usability criteria in a sample of 60-74–year-old Norwegians. The observed total energy intakes suggest that several of the participants underreported their intake during the completion of the dietary recall. Implementing web-based dietary assessment tools in older adults is feasible, but guidance and support might be needed to ensure valid completion.

Introduction

A healthy diet throughout the life course improves health and reduces the risk of disease [1]. The World Health Organization's Thirteenth General Programme of Work (GPW13) [2] emphasizes the encouragement of healthier life choices and behaviors including improvements in diet. The Norwegian National Action Plan for a Healthier Diet involves multiple objectives [3] in line with the United Nations (UN) Decade of Action on Nutrition 2016-2025 [4]. A key point is the need for research to enhance knowledge of the relation between diet and health [3].

Life expectancy is increasing, and the world's population is ageing [5]. Therefore, middle-aged and older adults are important target groups for monitoring food and nutrient intake as well as for dietary interventions to ensure healthy ageing. Information on food and nutrient intake can be collected from a variety of sources, including objective methods such as direct observation or analysis of biomarkers, and self-report methods, such as food frequency questionnaires (FFQs), 24-hour recalls, or food records. All methods have methodological limitations [6-8], including high costs and participant burden, and none are essentially optimal to accurately capture all elements of diet. Objective measurements, such as biomarkers, are not influenced by the participant's personal beliefs or perceptions. However, biomarkers do not cover all components of the diet and are affected by absorption and metabolism; hence, they are not always suitable or feasible and are dependent on the study context [7]. Self-report tools, such as FFQs assessing habitual intake, food records, or dietary recalls are noninvasive and usually feasible to use in a variety of research settings. However, self-reported measures are prone to large measurement errors occurring from the knowledge and beliefs of the study participants and rely on long- or short-term memory [7]. The 24-hour dietary recall is considered one of the most accurate self-report tools, as it collects information about diet within a short recent period; however, repeated measurements are required to capture the usual diet. New technology provides increased convenience for using the 24-hour dietary recall [6] and improves their usage in large-scale studies.

Measure Your Food On One Day (myfood24) is a self-administered web-based dietary data collection tool and analysis software developed [9] and validated [10] at the University of Leeds, the United Kingdom. The tool can be set up as a 24-hour dietary record or recall, for one or multiple days of prospective or retrospective recording. The myfood24 software can be used not only as a research instrument but also for teaching or clinical use. Region-adapted versions of myfood24 already exist for use in Australia, the Caribbean, Denmark, France, Germany, the Middle East, and Peru. The use of the same tool enables between-country comparisons. A Norwegian version of myfood24 [11] was developed and finalized in 2020 with further updates in 2021. The adaption from the original UK version consisted mainly of adapting the food database and images in the data bank to the Norwegian food cuisine, and the portion size images used in the Norwegian version has shown favorable properties in young adults [11].

The UK version of myfood24 has demonstrated reasonable response proportions for repeated 24-hour dietary recalls in 60-85–year-old individuals in the United Kingdom [12], and the German version has shown satisfactory usability in adults aged ≥65 years who were provided instructions and assistance during a study visit [13]. The usability of the Norwegian version has yet to be examined. There is an overall need for studies on the usability of web-based dietary data collection tools in the older adult population, as such data are scarce. It is of particular interest to examine the properties of the newly developed Norwegian version of myfood24, as this has not been examined before. Therefore, we aimed to examine the usability of a web-based 24-hour dietary recall tool, the Norwegian version of myfood24, in a sample of Norwegian older adult women and men who have been provided no prior tutoring or instructions.

Methods

Study Design and Participant Recruitment

We recruited participants by combining targeted convenience and snowball sampling to ensure an even distribution of sex, age (between 60-74 years), and educational level (elementary school, high school, or other, or higher or university education). Only individuals with an email address and access to a smartphone, tablet device, or computer were invited. We approached individuals accessible to the researchers through their personal networks (convenience sampling), and further invited individuals approached by already recruited individuals (snowball sampling). A total of 72 (convenience sample size) women and men aged 60-74 years (58% women) were invited to complete a single dietary recall. Invitations were sent in 3 rounds between May 12 and June 8, 2021, of which the first 2 groups received 1 reminder after 2 weeks (reminders were sent to all, as participation was anonymous).

Instruments

myfood24 24-Hour Dietary Recall

myfood24 [9,10] dietary recall can be completed on the internet using a smartphone, tablet device, or computer. The participants were asked to report the intake of food during the previous day (24 hours), a so-called 24-hour dietary recall. The myfood24 is partially based on a multiple pass method, including an optional first quick list, a foods search in addition to prompts both for forgotten foods (for commonly forgotten foods and those often combined with other foods) and a final review before submission of the recall [9]. Participants can search for and add food items from a detailed list of foods primarily based on the Norwegian Food Composition Table 2019 [14] and chose portion sizes from pictures, natural measures (eg, number of potatoes), or household measures (eg, number of spoons of sugar added), as well as in an open-text response option. The portion sizes are a combination of pictures from the original UK version and the newly developed Norwegian version [11]. Based on the reported food intake, the total energy, macronutrient, micronutrient, and food intake were calculated in myfood24 (except for information from the open-text response).

System Usability Scale

The System Usability Scale (SUS) is a postsession questionnaire [15] and one of the most commonly used usability scales to evaluate a variety of products and services including hardware, software, mobile devices, and websites. The user (study participant in this case) is asked to evaluate the system (myfood24 in this case) after a user test (completing the myfood24 web-based dietary recall in this case). The SUS has 10 items answered with a 5-point Likert scale from "Strongly disagree" to "Strongly agree," with alteration between positive and negative items to prevent response biases [16]. The 10 SUS items were as follows: (1) "I would like to use myfood24 again," (2) "I found myfood24 was unnecessarily complicated," (3) "I thought myfood24 was easy to use," (4) "I needed the support of a technical person to use myfood24," (5) "I found the various functions in myfood24 were well integrated," (6) "I thought there was too much inconsistency in myfood24," (7) "I imagine most people would be able to use myfood24 very quickly," (8) "I found myfood24 very cumbersome (awkward) to use," (9) "I felt very confident using myfood24," and (10) "I needed to learn a lot of things before I could get going with myfood24."

Data Collection

The invitees were sent an email with a weblink to myfood24 and an SMS text message to their mobile phone with information that the email had been sent. The email included a brief description of the aim of the study (ie, to study the usability of a web-based dietary data collection tool) and a common (ie, not individual) weblink to complete the 24-hour dietary recall and a questionnaire about usability and demographics. No detailed guidance about how to complete the myfood24 was provided; that is, no preparation or tutoring.

By clicking the weblink in the email, the participant was directed to myfood24 and asked to report all food and drink intake from the previous day. After completion of the 24-hour dietary recall (ie, by clicking "submit" in myfood24), the participant was automatically redirected to a separate external web-based questionnaire created using Nettskjema [17], a web-based questionnaire tool resource developed by the University of Oslo, Norway. The Nettskjema questionnaire included the 10 SUS items, 3 related to demographic characteristics—sex, age group (60-64, 65-69, or 70-74 years), and education level ("What is your highest level of education?" with response alternatives of "elementary school/high school/other" or "higher education/university")—along with 1 question about the device used to complete myfood24 ("smart phone," "iPad/tablet," or "PC/computer/Mac") and an open-ended question ("Do you have other comments?") where the participant could add supplemental information (as free text). Demographic data from Nettskjema could not be linked to the information about dietary intake from myfood24.

Ethics Approval and Consent to Participate

According to the Norwegian Health Research Act [18], analysis of anonymous data does not require ethical approval; therefore, this study is not evaluated by the Regional Committees for Medical and Health Research. In this study, true anonymity was ascertained through the following procedures: (1) the weblink

to the questionnaires (sent to the participants' personal emails) was a nonpersonal URL (ie, all participants received the same weblink with no possibility to identify individuals), (2) the demographic data collected were aggregated (ie, group-level data for age and education), and (3) demographic variables from Nettskjema could not be linked to the information about dietary intake from myfood24. The email to all participants included information about the study. Completion of the questionnaire was regarded as consent.

Data Analysis

The SUS score was calculated in accordance with Sauro 2011 [16]. The scoring is 0-4 for each item, where for odd items (positive) 1 is subtracted from the user response, and for even-numbered items (negative) 5 is subtracted from the user response, added up and multiplied by 2.5 (to convert the score to 0-100), with a score of 68 points or higher considered a "satisfactory product" [16]. We examined the mean (SD) SUS score and median (IQR) values as means as well as proportions within the satisfactory product cut-off limit in groups stratified by sex, age (60-69 and 70-74 years), and education. We used Stata (StataCorp LLC) to calculate study sample characteristics, including energy intakes, as well as total energy intake in kJ, including proportions with energy intakes outside standard cut-offs of <2100, >16,800 [19], and >8000 kJ (ie, 500, 4000, and 1900 kilocalories, respectively), mean and energy percentages (E%) for macronutrients, and mean SUS scores with 2-sample t test for differences between age groups (normal distribution observed with histograms). We used Excel (2016; Microsoft Corp) to present the response to each usability item as frequencies. Participants' free-text comments were grouped into themes.

Results

A total of 60 women and men participated (participation 83%, 57% women). Study sample characteristics and energy intakes are shown in Table 1. Half of the participants were in the oldest age group. Two-thirds reported having higher education, equally distributed by age. To complete myfood24, 41% (n=24) used a smartphone, 18% (n=11) used a tablet device, and 41% (n=24) used a computer. Two participants had not recorded any foods or beverages in myfood24. When excluding these, the mean and median total energy intake calculated from myfood24 were 5815 (SD 3093) kJ and 5896 (IQR 3815-7089) kJ, respectively, for the overall study sample. In total, 14% (n=8) had an energy intake of <2100 kJ, 17% (n=10) had an energy intake of >8000 kJ, and none had an energy intake of >16,800 kJ. The E% for all macronutrients were within the range of the national recommendations. In total, 10 participants recorded food items in the open-text option (eg, "home-made bun with spread for lunch" and "cloudberries with cream for dessert").

The response to each SUS item is presented in Figure 1, where the positive items (the 5 odd-numbered questions) and the negative items (the 5 even-numbered questions) have been grouped to enhance readability. The positive items were highly correlated, with responses 25%-28% above the middle category (agree or strongly agree), except for the item "I felt very confident using myfood24," where 40% (n=24) responded above

the middle category (agree or strongly agree). Two negative statements that stood out were "I needed support of a technical person to use myfood24®" where 72% (n=43) responded with "strongly disagree" and in total 85% (n=51) responded below the middle response category (disagree or strongly disagree), and "I needed to learn a lot of things before I could get going with myfood24®" where 68% (n=41) responded below the middle category (disagree or strongly disagree).

A total of 16 participants (27%) had SUS scores of ≥68 (ie, within the satisfactory product cut-off). Table 2 presents mean

SUS scores in women and men, low and high educational groups, and smartphone, tablet device, and computer users by age group. The mean and median overall SUS scores were 55.5 (SD 18.6) and 55.0 (IQR 41.3-70.0), respectively, with higher scores in the younger age group (*P*=.02). For each group stratified by sex, education, and device used, there was a tendency toward higher SUS scores in the younger compared to the older age group, but only significantly different among those with higher education.

Table 1. Participant characteristics and energy intakes for the overall sample (N=60) and by age group.

Characteristics	Total	Age group 60-69 years (n=30)	Age group 70-74 years (n=30)
Women, n (%)	34 (56.7)	19 (63.3)	15 (50.0)
Higher education, n (%)	44 (73.3)	21 (70.0)	23 (76.7)
Device, n (%)			
Smartphone	24 (40.7)	13 (43.3)	11 (37.9)
Tablet device	11 (18.6)	4 (13.3)	7 (24.1)
Computer	24 (40.7)	13 (43.3)	11 (37.9)
Total energy intake[a] (kJ), mean (SD)	5815 (3093)	—[b]	—
Carbohydrates (energy percentage), mean (SD)	41.7 (11.1)	—	—
Protein (energy percentage), mean (SD)	19.9 (6.0)	—	—
Fat (energy percentage), mean (SD)	32.9 (11.6)	—	—
Alcohol (energy percentage), mean (SD)	2.7 (5.9)	—	—
Fiber (energy percentage), mean (SD)	2.7 (1.3)	—	—

[a]Energy and nutrient intakes were only calculated for participants with a total energy intake of >0 kJ (n=58). One participant did not report the device used.

[b]—: not available.

Figure 1. Distribution (frequency) of the response to each System Usability Scale item after completion of myfood24.
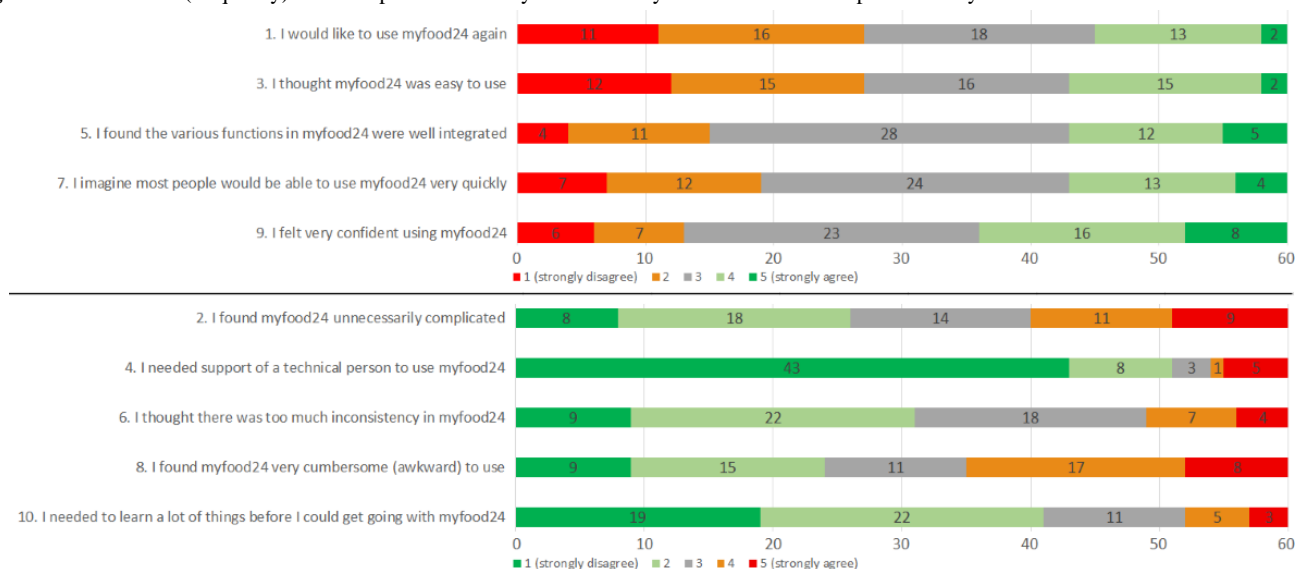
Table 2. Mean usability score[a] for the overall sample (N=60) and by age group.

Categories	Total, mean (SD)	Age group 60-69 years (n=30), mean (SD)	Age group 70-74 years (n=30), mean (SD)	P value[b]
Total	55.5 (18.6)	61.3 (17.6)	49.8 (18.0)	.02
Women	56.3 (20.5)	61.6 (18.4)	49.5 (21.7)	.09
Men	54.5 (16.0)	60.7 (17.0)	50.0 (14.1)	.09
Lower education	62.0 (14.4)	65.6 (15.0)	57.5 (13.2)	.28
Higher education	53.1 (19.5)	59.4 (18.6)	47.4 (18.9)	.04
Smartphone	52.5 (20.1)	61.9 (14.8)	55.5 (10.8)	.24
Tablet	55.9 (25.1)	68.1 (28.1)	48.9 (22.3)	.24
Computer	59.0 (13.3)	58.5 (17.6)	45.5 (21.5)	.12

[a]A System Usability Scale score of ≥68 indicates that the usability criteria were satisfied.

[b]A 2-sample *t* test was performed to assess differences between age groups.

In total, 18 participants had left comments in the free-text option. These comments (examples in parentheses) could thematically be grouped to (1) lack of categories of meals or foods ("Found nowhere to register supper"), (2) complexity ("Unnecessary complex at the same time as not nuanced enough to catch important differences between foods"), (3) wording ("Why is the word 'generic' used?"), or (4) technical problems ("Two times nearly finished, got kicked out").

Discussion

Principal Findings

In this study of the usability of a new web-based dietary assessment tool among older adult Norwegian women and men who were provided no guidance or support, we found the overall usability to be less than satisfactory and the overall total energy intake to imply error (underreporting) in completion of the dietary recall. In total, 17% of participants had energy intakes above 8000 kJ, and 27% had usability scores within the range indicating a satisfactory product.

The E% for all macronutrients were within the recommendations from the Norwegian Health Directorate [20], but in the total sample, the calculated mean total energy intake was only 60% of the daily total energy intake of an average inactive adult. This could imply that, for most participants, a variety of foods were recorded, but some foods (or meals) were probably omitted from the recording, or the amounts reported were incorrect. A traditional 4-meal Norwegian diet consists of breakfast (typically bread with spread or cereal with milk or yogurt), lunch (typically sandwiches), dinner (typically a dish with vegetables and meat or fish), and a light supper (typically similar foods as in breakfast or lunch), while myfood24 is based on 3 main meals, and other meals are classified as snacks. Based on the qualitative feedback provided in the open-text fields, some of the participants seem to struggle with this difference. Further, a total of 16% recorded food in the open-text response (instead of finding an alternative in the foods list), and these foods were omitted from the calculation of energy intake. Including these items in the calculations would have increased the mean energy intake. Misreporting is, however, common in all self-administered tools for all age groups, and for dietary assessment tools most often in the direction of underreporting, leading to incorrectly low estimates of total energy intakes.

Our results deviate considerably from the findings from a usability study of the German version of myfood24 [13], who found a median SUS score above the good product cut-off (median 74, IQR 70-83) among older adults aged ≥65 years, as well as total energy intake within the normal range, implying that the study participants completed myfood24 correctly. A potential explanation for this difference could be that the data collection in the German study was performed during a study center visit, where the participants received instructions on how to complete myfood24 as well as on-site assistance [13]. Further, the participants completed 3-4 consecutive recalls (ie, probably resulting in the participant being acquainted with the tool), 11% had previous experience in completing 24-hour dietary recalls, and 21% had a nutritional or food science background [13], of which both are expected to increase the perception of usability. Furthermore, the German version of myfood24 contains a higher number of foods than the Norwegian version, including a large number of branded foods. This can make it easier to determine the exact food consumed, thus increasing usability. However, there is a need for balance in the number of foods listed. If the list of foods is regarded as too extensive by the user, this can decrease usability.

During the development of the original UK version of myfood24, a small sample of adults aged above 65 years was included to report expectations from myfood24 [10]. The sample's median SUS score was poor (29, IQR 63); however, this sample consisted of only 4 participants [9]. Thus, the findings could have resulted from chance events. A review of another web-based 24-hour recall tool, the Automated Self-Administered 24-hour (ASA24) Dietary Assessment Tool used in Canadian populations found that older adults encountered technical challenges and emphasized the need for tailored support [21]. However, the need for technical assistance was not limited to older adults, and although older adults had a lower technical troubleshooting capacity, they simultaneously were willing to spend more time troubleshooting than younger adults [21]. There is strong reason to believe that with increasing ownership of smartphones and tablets as well as preference to

use digital technologies in everyday life, the feasibility of using web-based dietary assessment tools will also increase in the older adult population [6].

Nonetheless, digital illiteracy is a concern for self-administered web-based assessment tools, particularly among older adults. In a feasibility study of myfood24 among 60-85–year-old individuals in the United Kingdom [12], those completing multiple recalls reported higher self-confidence with technology and had a higher technology readiness score than those who did not complete any recalls. Notably, the participants in the UK study were recruited from an ongoing cohort study of cognitively healthy older adults [12], which could introduce a risk of selection bias.

Even though the overall SUS score in our study was lower than that considered satisfactory, we do believe that myfood24 is a useful tool to measure dietary intake. The participants in our study were solely provided information about the aim of the study, with no guidance or support in completing the recall. We believe that additional tailored instruction on how to complete the recall could have improved the overall perception of usability considerably and led to less underreporting. Moreover, usability questions highlighting the need for technical support or training revealed that technical challenges did not constitute a major barrier. Thus, we believe that technical difficulty was not the main contributor to the observed SUS score. In telephone surveys by Statistics Norway, 70% of 75-79–year-old individuals reported daily use of the internet in 2021 compared to 40% in 2015 [22], implying that digital illiteracy is decreasing in Norwegian older adults. Therefore, we believe that the feasibility of using web-based assessment tools in older adults will increase substantially over the next few years. More work is needed in adapting the Norwegian version of myfood24 to older adults, such as explaining how to record supper. Based on the qualitative feedback from comments provided by the participants in this study, we also believe that adding more foods and more synonyms to the foods list would improve the user experience. Previous usability studies of self-administered

dietary assessment apps among older adults, include apps with voice-added reporting [23] or apps where health personnel can complete the reporting on behalf of the respondent [24]. For older adults with impairments, such modifications are helpful. Other apps are developed for clinical use by health personnel only. In addition to being a self-administrative tool for research or teaching purposes, myfood24 provides a separate health care solution for health personnel to set patients' personalized nutrient targets and track their progress. Thus, modification to fit the context and the need of the individual is possible when using myfood24.

Strengths and Limitations

The study was set up to be anonymous to protect participant privacy; therefore, the participants received unidentifiable weblinks. Thus, a major study limitation is that we could not link SUS scores and dietary intakes on an individual level, to examine the correlation between intake and SUS scores. Further, we collected no information regarding whether the participants were comfortable with using technology. We invited only women and men with an email address and access to a device to complete the recall, and we did not examine reasons for nonparticipation. People without email addresses or those who did not participate could be less digitally literate. Thus, lower SUS scores could be anticipated among nonparticipants. A strength of the study is testing of the tool in a mixed group of older adults (ie, both sexes, different ages, and educational levels).

Conclusions

In this population of older adult Norwegian women and men, the overall usability score for a web-based 24-hour dietary recall assessment tool, myfood24, was lower than what is considered satisfactory according to standard usability criteria. Implementing myfood24 in older adults is feasible, but additional guidance and support might be needed to ensure valid completion, and more work is needed for adapting myfood24 for use among older adult Norwegians.

Conflicts of Interest

The authors declare that they have no competing interests. ACM was the principal investigator in developing the Norwegian version of myfood24. She holds an unpaid advisory position in Dietary Assessment Ltd and receives no personal financial benefits from others using the tool.

References

1. Healthy diet. World Health Organization. 2020 Apr 29. URL: https://www.who.int/news-room/fact-sheets/detail/healthy-diet [accessed 2021-11-01]
2. Thirteenth General Programme of Work (GPW13) Methods for impact measurement. World Health Organization. 2020 Nov 09. URL: https://www.who.int/publications/m/item/thirteenth-general-programme-of-work-(gpw13)-methods-for-impact-measurement [accessed 2021-11-01]
3. Norwegian National Action Plan for a Healthier Diet – an outline. Norwegian Ministry of Health and Care Services. URL: https://www.regjeringen.no/contentassets/fab53cd681b247bfa8c03a3767c75e66/norwegian_national_action_plan_for_a_healthier_diet_an_outline.pdf [accessed 2021-11-01]
4. United Nations decade of action on nutrition 2016-2025. United Nations. 2016. URL: https://www.un.org/nutrition/ [accessed 2021-11-01]
5. Partridge L, Deelen J, Slagboom PE. Facing up to the global challenges of ageing. Nature 2018 Sep;561(7721):45-56. [doi: 10.1038/s41586-018-0457-8] [Medline: 30185958]

6. Amoutzopoulos B, Steer T, Roberts C, Cade JE, Boushey CJ, Collins CE, et al. Traditional methods new technologies - dilemmas for dietary assessment in large-scale nutrition surveys and studies: a report following an international panel discussion at the 9th International Conference on Diet and Activity Methods (ICDAM9), Brisbane, 3 September 2015. J Nutr Sci 2018;7:e11 [FREE Full text] [doi: 10.1017/jns.2018.4] [Medline: 29686860]

7. Shim J, Oh K, Kim HC. Dietary assessment methods in epidemiologic studies. Epidemiol Health 2014;36:e2014009 [FREE Full text] [doi: 10.4178/epih/e2014009] [Medline: 25078382]

8. Cade JE. Measuring diet in the 21st century: use of new technologies. Proc Nutr Soc 2017 Aug;76(3):276-282. [doi: 10.1017/S0029665116002883] [Medline: 27976605]

9. Carter MC, Albar SA, Morris MA, Mulla UZ, Hancock N, Evans CE, myfood24 Consortium. Development of a UK online 24-h dietary assessment tool: myfood24. Nutrients 2015 May 27;7(6):4016-4032 [FREE Full text] [doi: 10.3390/nu7064016] [Medline: 26024292]

10. Wark PA, Hardie LJ, Frost GS, Alwan NA, Carter M, Elliott P, et al. Validity of an online 24-h recall tool (myfood24) for dietary assessment in population studies: comparison with biomarkers and standard interviews. BMC Med 2018 Aug 09;16(1):136 [FREE Full text] [doi: 10.1186/s12916-018-1113-8] [Medline: 30089491]

11. Salvesen L, Engeset D, Øverby NC, Medin AC. Development and evaluation of image-series for portion size estimation in dietary assessment among adults. J Nutr Sci 2021;10:e3 [FREE Full text] [doi: 10.1017/jns.2020.58] [Medline: 33889386]

12. Ward HA, McLellan H, Udeh-Momoh C, Giannakopoulou P, Robb C, Wark PA, et al. Use of online dietary recalls among older UK adults: a feasibility study of an online dietary assessment tool. Nutrients 2019 Jun 27;11(7):1451 [FREE Full text] [doi: 10.3390/nu11071451] [Medline: 31252526]

13. Koch SAJ, Conrad J, Hierath L, Hancock N, Beer S, Cade JE, et al. Adaptation and evaluation of Myfood24-Germany: a web-based self-administered 24-h dietary recall for the German adult population. Nutrients 2020 Jan 06;12(1):160 [FREE Full text] [doi: 10.3390/nu12010160] [Medline: 31935885]

14. Norwegian Food Composition Database 2019. Norwegian Food Safety Authority. URL: https://www.matvaretabellen.no/ [accessed 2021-11-01]

15. Brooke J. SUS: A 'Quick and Dirty' Usability Scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. London: Taylor and Francis; 1996:189-194.

16. Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. Denver, CO: Measuring Usability LLC; 2011.

17. UIO. Nettskjema. University of Oslo. Oslo URL: https://nettskjema.no/ [accessed 2022-02-24]

18. Act on Medical and Health Research (Health Research Act). Lovdata. 2008. URL: https://lovdata.no/dokument/NL/lov/2008-06-20-44#KAPITTEL_1/ [accessed 2022-03-08]

19. Willett W. Nutritional Epidemiology (3rd Edition). New York, NY: Oxford University Press; 2013.

20. Energi. Matportalen.no. URL: https://www.matportalen.no/kosthold_og_helse/tema/naringsstoffer/energi [accessed 2021-11-01]

21. Kirkpatrick SI, Gilsing AM, Hobin E, Solbak NM, Wallace A, Haines J, et al. Lessons from studies to evaluate an online 24-hour recall for use with children and adults in Canada. Nutrients 2017 Jan 31;9(2):100 [FREE Full text] [doi: 10.3390/nu9020100] [Medline: 28146125]

22. Bruk av IKT i husholdningene. Statistisk sentralbyrå. URL: https://www.ssb.no/teknologi-og-innovasjon/informasjons-og-kommunikasjonsteknologi-ikt/statistikk/bruk-av-ikt-i-husholdningene [accessed 2021-11-01]

23. Liu Y, Chen C, Lin Y, Chen H, Irianti D, Jen T, et al. Design and usability evaluation of mobile voice-added food reporting for elderly people: randomized controlled trial. JMIR Mhealth Uhealth 2020 Sep 28;8(9):e20317 [FREE Full text] [doi: 10.2196/20317] [Medline: 32985999]

24. Coleone JD, Bellei EA, Roman MK, Kirsten VR, Bertoletti De Marchi AC. Dietary intake and health status of elderly patients with type 2 diabetes mellitus: cross-sectional study using a mobile app in primary care. JMIR Form Res 2021 Aug 27;5(8):e27454 [FREE Full text] [doi: 10.2196/27454] [Medline: 34448711]

Abbreviations

ASA24: Automated Self-Administered 24-hour
E%: energy percentage
FFQ: food frequency questionnaire
GPW13: World Health Organization's Thirteenth General Programme of Work
myfood24: Measure Your Food On One Day
SUS: System Usability Scale
UN: United Nations

XSL•FO
RenderX

Original Paper

Ascertaining Medication Use and Patient-Reported Outcomes via an App and Exploring Gamification in Patients With Multiple Sclerosis Treated With Interferon β-1b: Observational Study

Volker Limmroth[1], MD; Kirsten Bayer-Gersmann[2], BEng; Christian Mueller[3], PhD; Markus Schürks[3], MSc, MD

[1]Clinic for Neurology and Palliative Medicine, Municipal Hospital Köln-Merheim, Cologne, Germany

[2]Institut Dr. Schauerte, Munich, Germany

[3]Bayer Vital GmbH, Leverkusen, Germany

Corresponding Author:
Markus Schürks, MSc, MD
Bayer Vital GmbH
Medical Affairs, Neurology, Immunology & Ophthalmology
Leverkusen, 51368
Germany
Phone: 49 214 30 51455
Email: markus.schuerks@bayer.com

Related Article:

This is a corrected version. See correction statement: https://formative.jmir.org/2022/3/e38002

Abstract

Background: The BETACONNECT autoinjector and myBETAapp app were designed to support patients with multiple sclerosis receiving interferon β-1b and are an ideal platform for digital observational studies. A recent pilot study in Germany demonstrated the feasibility of using the app to recruit patients, obtain informed consent, and evaluate medication-taking behavior over 6 months.

Objective: This study aims to describe medication-taking behavior for 1 year in patients with multiple sclerosis receiving interferon β-1b based on data collected from the app and to provide information on patient-reported outcomes (PROs). The optional use of the cognitive training tool PEAK (Peak, formerly Brainbow Ltd) is included to test the feasibility of gamification in this setting.

Methods: A prospective and retrospective, exploratory, digital, observational cohort study was conducted among users of the app in Germany. Invitations to participate were sent to patients' apps between February and May 2019. Participants provided electronic informed consent. Injection-related data from consenting patients' devices were collected prospectively for 1 year following the consent date and retrospectively for ≤1 year from the first day of use (if historical data were available). Participants also completed three electronic PRO instruments every 3 months: the EuroQol 5-Dimension, 5-Level questionnaire (EQ-5D-5L); the Treatment Satisfaction Questionnaire for Medication (TSQM; version II); and a questionnaire on satisfaction with treatment support (on a server accessed via an emailed hyperlink). All patients were offered optional access to the professional version of PEAK.

Results: Of 1778 registered app accounts (May 2019), 79 patients (4.44%) provided informed consent; 62 (3.49%) were eligible for inclusion in the prospective analysis, of whom, 60 (97%) also had retrospective data. The mean age of the 62 participants was 43.2 (SD 11.5) years and 41 (66%) were women. Compliance over the 1-year prospective observational period (primary end point) was high (median 98.9%, IQR 94.3%-100%) and similar among men and women. Persistence and adherence (coprimary end points) decreased from 85% (53/62) and 74% (46/62), respectively, at 6 months to 76% (47/62) and 65% (40/62), respectively, at 12 months; both were higher in men than in women. A retrospective analysis showed similar patterns. The PRO questionnaires were answered by 79% (49/62) of the participants at baseline and 50% (31/62) of them at month 12. Women had more severe problems in some EQ-5D-5L dimensions (mobility, usual activities, and pain/discomfort) and lower median convenience scores on the TSQM (version II) than men. At month 12, 84% (26/31) of the patients were satisfied or very satisfied with the app. PEAK was used by 67% (14/21) of men and 49% (20/41) of women.

Conclusions: This study showed high compliance and decreasing persistence and adherence over 1 year and demonstrated the feasibility of including remotely completed electronic PRO instruments in digital observational studies.

Introduction

Background

Digital clinical studies use technology to support recruitment and retention, data collection, and analytics [1], giving them several important advantages over conventional clinical studies. For example, slow recruitment is a common problem in clinical studies [2]. Digital studies allow rapid recruitment of patients outside of routine clinic visits (eg, by directly approaching them via their smartphones), with a corresponding reduction in study duration. In conventional clinical trials, the number of enrolled patients and number of clinic visits have been identified as key drivers of cost [3]. The inconvenience of making additional clinic visits is also a barrier to participation in clinical trials [4]. Digital studies avoid the need for clinic visits and allow enrollment of large numbers of patients at little extra cost once the digital study platform has been established.

In Germany, many patients with multiple sclerosis (MS) treated with interferon β-1b use the BETACONNECT autoinjector, which automatically records injections. An app (myBETAapp) allows patients who are being treated with interferon β-1b to document injection data manually or by automatic transfer from the autoinjector. This system is designed to support patients with MS, and it provides an ideal platform for digital observational studies. A recent pilot study in Germany demonstrated the feasibility of performing an observational study with digital recruitment, consent, and data collection performed via the app [5]. The pilot study results supported the extension of the digital observational approach to more comprehensive studies investigating clinical and patient-reported outcomes (PROs) over longer periods [5,6].

Objectives

The aim of this digital observational study is to investigate medication-taking behavior in patients with MS receiving interferon β-1b for 1 year and to provide additional information on PROs. As cognitive impairment is common in MS [7], optional use of the cognitive training tool PEAK was included to test the feasibility of gamification in this setting and to incentivize patients to remain committed to the study.

Specifically, the primary objective is to describe medication-taking behavior prospectively for 1 year in patients with MS treated with interferon β-1b based on data collected from the app.

The secondary objectives are as follows: to investigate past medication-taking behavior (since first documentation in the app) before participation in the study; to evaluate health-related quality of life (HRQoL), treatment satisfaction, and satisfaction

with treatment support prospectively for 1 year; to investigate HRQoL, treatment satisfaction, and satisfaction with treatment support at baseline in patients who were adherent or nonadherent as well as those who were persistent or nonpersistent over 1 year; to assess the relationship of HRQoL, treatment satisfaction, and satisfaction with treatment support at baseline with medication-taking behavior at 6 months and 1 year; and to evaluate how frequently and for how long the smartphone-based cognitive training tool PEAK was used.

Methods

Study Design and Patients

This was a prospective and retrospective, noninterventional, observational cohort study with the structure of a registry (ClinicalTrials.gov NCT03808142). Adult patients with MS treated with interferon β-1b, in Germany, were eligible to participate in the study if they were new or existing users of the app and provided electronic informed consent.

Study Conduct and Ethics Approval

During a 3-month enrollment period (February 20, 2019, to May 19, 2019), all patients with an active account for the app were invited to participate in the study. After 2 months, invitations were sent to all new users of the app (patients who started using the app during the enrollment period). In both cases, the invitations consisted of push messages sent to the patients' apps. Patients who did not respond to the initial invitation were sent a reminder after 2 weeks.

Patients who expressed interest in the study (by pressing a button within the app) were presented with a detailed informed consent form. The form provided study information in a sequential manner (background, aim of study, study design and data use, data privacy including how to withdraw consent, and contact information for the database hosts in case of questions), and the patients were required to confirm after each sequence that they understood the information and wished to participate. The form also clarified that prospectively recorded data and, among existing users of the app, injection-related data recorded in the past would be analyzed. It also included text to advise patients to report any side effects or possible side effects to their physicians, nurses, or Bayer, and to report any technical product complaints to Bayer. Only patients who consented to all the steps were able to participate in the study.

The study protocol was approved by the ethics committee of the Nordrhein Medical Chamber (number 2018381). In this observational study, interferon β-1b was prescribed in routine clinical practice by the treating physician in accordance with the terms of the marketing authorization, and the patient was

offered support from the BETAPLUS Patient Support and Disease Management Program (PSDMP) as normal. The prescription of interferon β-1b and provision of the BETAPLUS PSDMP were clearly separated from the decision to include the patient in the study. No additional diagnostic or monitoring process was required for enrollment or during the study. Each patient could refuse to participate further or withdraw from the study at any time and without giving a reason. After withdrawal of the patient from the study, data from that patient were not used for further analyses. Each patient was assigned a unique central patient identification code, which was used only for study purposes.

Data Collection

Demographic data (age and gender) were recorded in the app as part of the registration process. Injection-related data (date and time of injection, injection speed, and injection depth) were automatically recorded in the BETACONNECT for each injection and could be transferred to the app if the patient wished to do so. Patients could also manually record the injection data in the app. Demographic and injection-related data in the app were automatically transferred to an external database (hosted by TWT Digital Health GmbH) whenever the mobile device was connected to the web.

In addition, the patients completed three PRO instruments every 3 months from the time of informed consent to the end of observation: the EuroQol 5-Dimension, 5-Level questionnaire (EQ-5D-5L); the Treatment Satisfaction Questionnaire for Medication (TSQM; version II) [8,9]; and a questionnaire on satisfaction with the BETAPLUS PSDMP, the BETACONNECT, and the app (service questionnaire; Multimedia Appendix 1). For this purpose, each patient was provided with an individual log-in hyperlink (newly created for each data capture time point) via email to enter an electronic data capture server hosted by the contract research organization, Institut Dr. Schauerte (IDS). Hence, the participants were able to complete the questionnaires remotely. After the end of the observation period, participants received vouchers for completing each questionnaire.

Each patient was followed up for 1 year after providing informed consent (prospective analysis). For patients who had already started using the app before providing informed consent, prospectively collected injection-related data (BETACONNECT or manual records) were also available from the past. Among those patients, medication-taking behavior was also analyzed from the time of the first documented injection to 1 year later or the date of consent, whichever was earlier (retrospective analysis; Figure 1).

Figure 1. Examples of retrospective and prospective analysis periods for patients with a first documented injection at different intervals before the date of consent.
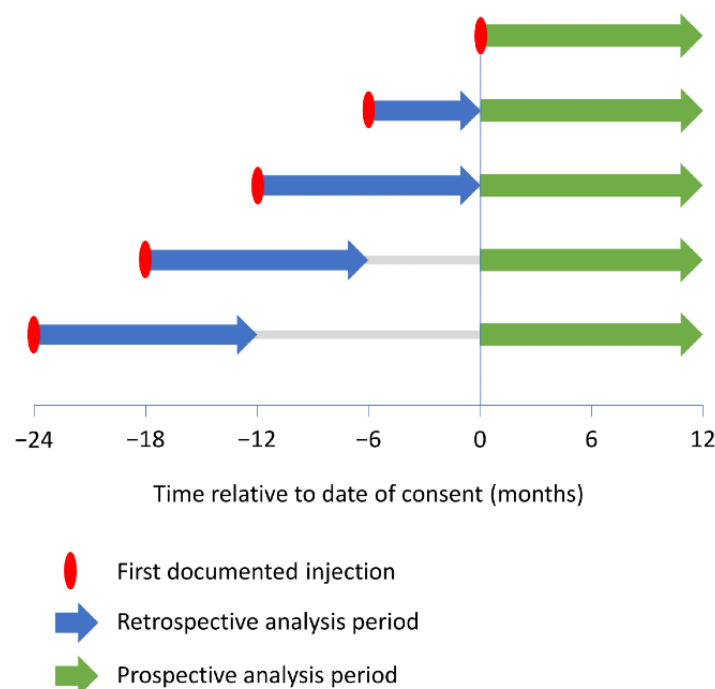
All patients were initially offered access to the professional version of the mobile-based cognitive training tool PEAK, developed and hosted by Peak (formerly Brainbow Ltd); this offer was renewed with each request to complete the PRO questionnaires (quarterly emailing of log-in hyperlinks). The patients were able to choose from all available games within PEAK, covering 7 cognitive areas, for the duration of the study. The decision to use the training tool and game selection was

purely voluntary and was not linked to treatment, study participation, or completion of the PRO instruments. All data generated by the patients while using PEAK were stored in a database hosted by Peak. The patients received study-specific log-in data, which identified them as participants of the study in the database, without personal identification.

Log-in details for the questionnaires on the IDS electronic data capture server and for PEAK were allocated to participants by

TWT Digital Health GmbH. Hence, only TWT Digital Health GmbH had the key to match data from Peak, IDS, and their own database. Data sets coded by log-in identifiers were transferred from Peak and TWT Digital Health GmbH to IDS, who merged these into their own data set. No investigator was involved in the data collection process.

Before the analysis, redundancies in the injection data were automatically corrected as described previously [5].

Outcome Measures

The primary end point was prospectively assessed compliance (the percentage of expected doses actually injected), based on injections recorded in the TWT Digital Health GmbH database from the date of informed consent to the end of the observation period.

The coprimary end points were persistence (the number of patients still using interferon β-1b at the end of the observation period), missed injections (expressed as a proportion of expected injections), and adherence (the number of patients who were both persistent and ≥80% compliant; a threshold of 80% is commonly used in studies of adherence in MS [10]). Patients without injection recordings for >2 weeks were regarded as nonpersistent.

Compliance, persistence, missed injections, and adherence were also analyzed retrospectively (from the first injection recorded in the app until 1 year later or the date of informed consent, whichever was earlier) as secondary end points among participants who had registered and used the app before enrolling in this study.

Other secondary end points included HRQoL assessed using the EQ-5D-5L, treatment satisfaction assessed using the TSQM (version II) [8,9], satisfaction with treatment support assessed using the service questionnaire (Multimedia Appendix 1), and use of the PEAK cognitive training tool. Frequency (trainings/week) and duration (days) of PEAK use by study participants and cumulative time played per game (minutes) were analyzed. For the analysis of the frequency of PEAK use and cumulative time played per game, only the 10 most frequently used games were considered.

Statistical Analysis

Statistical analyses were exploratory and descriptive, using summary statistics for categorical and quantitative (continuous) data. Associations of persistence and adherence with baseline HRQoL, treatment satisfaction, and satisfaction with support were evaluated using univariate logistic regression analysis, with $P<.05$ considered significant. All statistical analyses were performed using the software package SAS 9.4 (SAS Institute Inc).

Assuming an estimated SD of 15 for the mean compliance (as observed in the pilot study [5]), a sample size of 100 would produce a 2-sided 95% CI with a half-width of 3.0.

Results

Participants

Of the 1778 registered accounts for the app (May 2019), 79 (4.44%) patients provided informed consent to participate in the study. Of the 79 patients, 10 (13%) patients withdrew their consent, 6 (8%) patients were excluded because they were enrolled after the end of the registration period, and 1 (1%) patient had no injection or questionnaire data. Thus, 78% (62/79) of patients had injection-related data and were valid for inclusion in the prospective analysis. Of the 62 patients, 2 (3%) had no retrospective injection data and were therefore excluded from the retrospective analysis.

The mean age of the 62 participants was 43.2 (SD 11.5) years; 41 (66%) were women, and 21 (34%) were men. The distribution of age groups in women and men is presented in Table 1.

Table 1. Study participants according to age group and gender.

Age group (years)	Participants, n (%)		
	Total (N=62)	Female (n=41)	Male (n=21)
<30	9 (15)	8 (20)	1 (5)
30 to <40	15 (24)	9 (22)	6 (29)
40 to <50	19 (31)	11 (27)	8 (38)
50 to <60	13 (21)	9 (22)	4 (19)
≥60	6 (10)	4 (10)	2 (10)

Data Collected

Of the 62 patients, 14 (23%) used the BETACONNECT autoinjector to administer and record injections throughout the prospective observation period. For 19% (12/62) of the patients, only manual records of injections were available. Alternating recording methods (autoinjector or manual) were documented in 56% (35/62) of the patients. Switching from one mode of injection to the other (autoinjector to manual or vice versa) was documented in 2% (1/62) of the patients.

Use of the BETACONNECT to record injections was more common among men (8/21, 38%) than among women (6/41, 15%), whereas only manual recording of injections was more common among women (10/41, 24%) than among men (2/21, 10%). The injection location data were recorded for all 62 patients.

For the prospective analysis (N=62), 7 (11%) patients had <6 months of injection data and 55 (89%) patients had ≥6 months of injection data, with 48 of the 55 patients (87%) having ≥12

months of injection data. For the retrospective analysis (n=60), 8 (13%), 52 (87%), and 44 (73%) patients had <6, ≥6, and ≥12 months of injection data, respectively.

Of the 62 patients, 53 (85%) answered at least one questionnaire, and 34 (55%) used PEAK at least once. The EQ-5D-5L, TSQM (version II), and service questionnaire were each answered by 49 (79%) of the 62 patients at baseline and by 36 (58%), 35 (56%), 32 (52%), and 31 (50%) patients at months 3, 6, 9, and 12, respectively.

Medication-Taking Behavior

In the prospective analysis, persistence was 85% (53/62) at 6 months and decreased to 76% (47/62) at 12 months after the date of informed consent (Table 2). Persistence was higher in men than in women at both 6 months (20/21, 95% vs 33/41,

80%) and 12 months (19/21, 90% vs 28/41, 68%; Table 2). Compliance (the primary end point, based on injection data up to 6 months and up to 12 months in patients with injection data available for those time periods) was high at 6 months (mean 93.6%, SD 13.8%; median 100%, IQR 95.6%-101.1%) and at 12 months (mean 92.6%, SD 14.1%; median 98.9%, IQR 94.3%-100%), and was similar in men and women at both time points (Table 2 and Figure 2Ai). Mean and median compliance were high in all age groups (Figure 2Ai). In total, 69% (43/62) and 56% (35/62) of the patients had missed only 0% to 5% of their expected injections at 6 and 12 months, respectively (Figure 2Bi). Adherence was markedly higher in men than in women (Table 2 and Figure 2Ci). At 6 months, adherence showed no clear difference across age groups, whereas at 12 months, it was lowest in the youngest age group (Figure 2Ci).

Table 2. Compliance, persistence, and adherence analyzed prospectively and retrospectively.

Characteristics	6 months			12 months		
	Total	Female	Male	Total	Female	Male
Prospective analysis						
Patients, n	62	41	21	62	41	21
Persistence, n (%)						
Yes	53 (85)	33 (80)	20 (95)	47 (76)	28 (68)	19 (90)
No	9 (15)	8 (20)	1 (5)	15 (24)	13 (32)	2 (10)
Compliance (%)[a]						
Values, mean (SD)	93.6 (13.8)	91.2 (15.5)	97.7 (9)	92.6 (14.1)	90.6 (15.7)	95.4 (11.4)
Values, median (IQR)	100 (95.6-101.1)	97.8 (92.2-101.1)	101.1 (99.4-101.1)	98.9 (94.3-100)	97.3 (92.3-100)	99.7 (96.4-100)
Values, range	47.8-103.3	47.8-101.1	64.4-103.3	41.5-101.1	41.5-100.5	53-101.1
Adherence, n (%)						
Yes	46 (74)	27 (66)	19 (90)	40 (65)	22 (54)	18 (86)
No	16 (26)	14 (34)	2 (10)	22 (35)	19 (46)	3 (14)
Retrospective analysis						
Patients, n	60	39	21	60	39	21
Persistence, n (%)						
Yes	50 (83)	31 (79)	19 (90)	42 (70)	24 (62)	18 (86)
No	10 (17)	8 (21)	2 (10)	18 (30)	15 (38)	3 (14)
Compliance (%)[a]						
Values, mean (SD)	96.7 (13)	95.7 (15.5)	98.4 (6.7)	96.8 (8.7)	95.8 (10.7)	98.2 (4.3)
Values, median (IQR)	101.1 (98.9-101.1)	101.1 (98.9-101.1)	100 (100-101.1)	100 (97.3-100)	100 (97.3-100)	99.7 (97.3-100)
Values, range	27.8-107.8	27.8-107.8	77.8-107.8	52.5-104.9	52.5-103.3	88-104.9
Adherence, n (%)						
Yes	48 (80)	30 (77)	18 (86)	41 (68)	23 (59)	18 (86)
No	12 (20)	9 (23)	3 (14)	19 (32)	16 (41)	3 (14)

[a]On the basis of the injection data up to 6 months and up to 12 months in patients with injection data available for the periods (prospective analysis: n=55 at 6 months and n=48 at 12 months; retrospective analysis: n=52 at 6 months and n=44 at 12 months).

Figure 2. (A) Compliance overall and by gender and age group, (B) patients categorized by percentage of interferon β-1b injections missed (based on an expected frequency of 1 injection every other day), and (C) adherence overall and by gender and age group, analyzed (i) prospectively and (ii) retrospectively. Compliance at 6 and 12 months was assessed in patients with ≥6 and ≥12 months of injection-related data, respectively. In the box and whisker plots, the colored bars indicate median and IQR, the whiskers indicate minimum and maximum values, and the white circles indicate the mean.
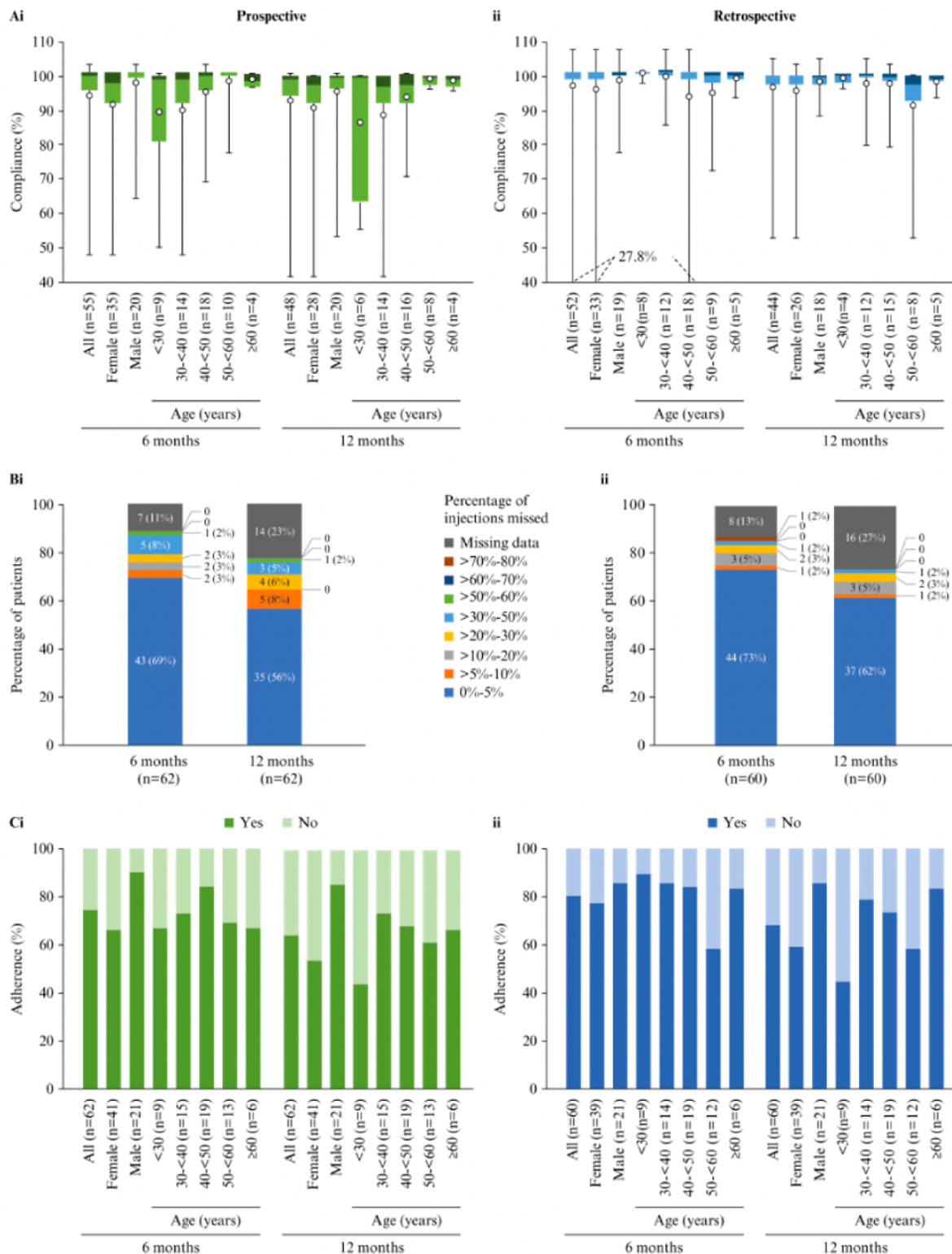
In the retrospective analysis, persistence was 83% (50/60) at 6 months and markedly lower, 70% (42/60), at 12 months. It was higher in men than in women at 6 months (19/21, 90% vs 31/39, 79%) and 12 months (18/21, 86% vs 24/39, 62%) (Table 2). Mean and median compliance (based on injection data up to 6 months and up to 12 months in patients with injection data available for those time periods) were high in both men and women (Table 2 and Figure 2Aii) and in all age groups at both time points (Figure 2Aii). In total, 73% (44/60) and 62% (37/60) of the patients had missed only 0% to 5% of their expected injections at 6 and 12 months, respectively (Figure 2Bii). Adherence was higher in men than in women at both time points (Table 2 and Figure 2Cii). In the analysis by age group, those aged 50 to <60 years showed the lowest adherence at 6 months,

whereas those aged <30 years showed the lowest adherence at 12 months (Figure 2Cii).

HRQoL and Satisfaction With Treatment and Support

In the EQ-5D-5L analysis, women reported more severe problems with mobility, usual activities, and pain/discomfort than men (Figure 3). None of the dimensions showed a consistent difference across age groups. The median (IQR) EQ-5D-5L index was 0.9 (0.8-1.0) at baseline and 0.9 (0.9-1.0) at 6 and 12 months. The median (IQR) visual analogue scale (VAS) was 81.0 (65.0-90.0) at baseline, 82.0 (62.0-90.0) at 6 months, and 80.0 (66.0-88.0) at 12 months. The EQ-5D-5L index and VAS showed no consistent differences between men and women or across age groups.

Figure 3. Patient-reported assessment of (A) mobility, (B) self-care, (C) usual activities, (D) pain/discomfort, and (E) anxiety/depression at (i) baseline, (ii) 6 months, and (iii) 12 months in the EuroQol 5-Dimension, 5-Level questionnaire (EQ-5D-5L).
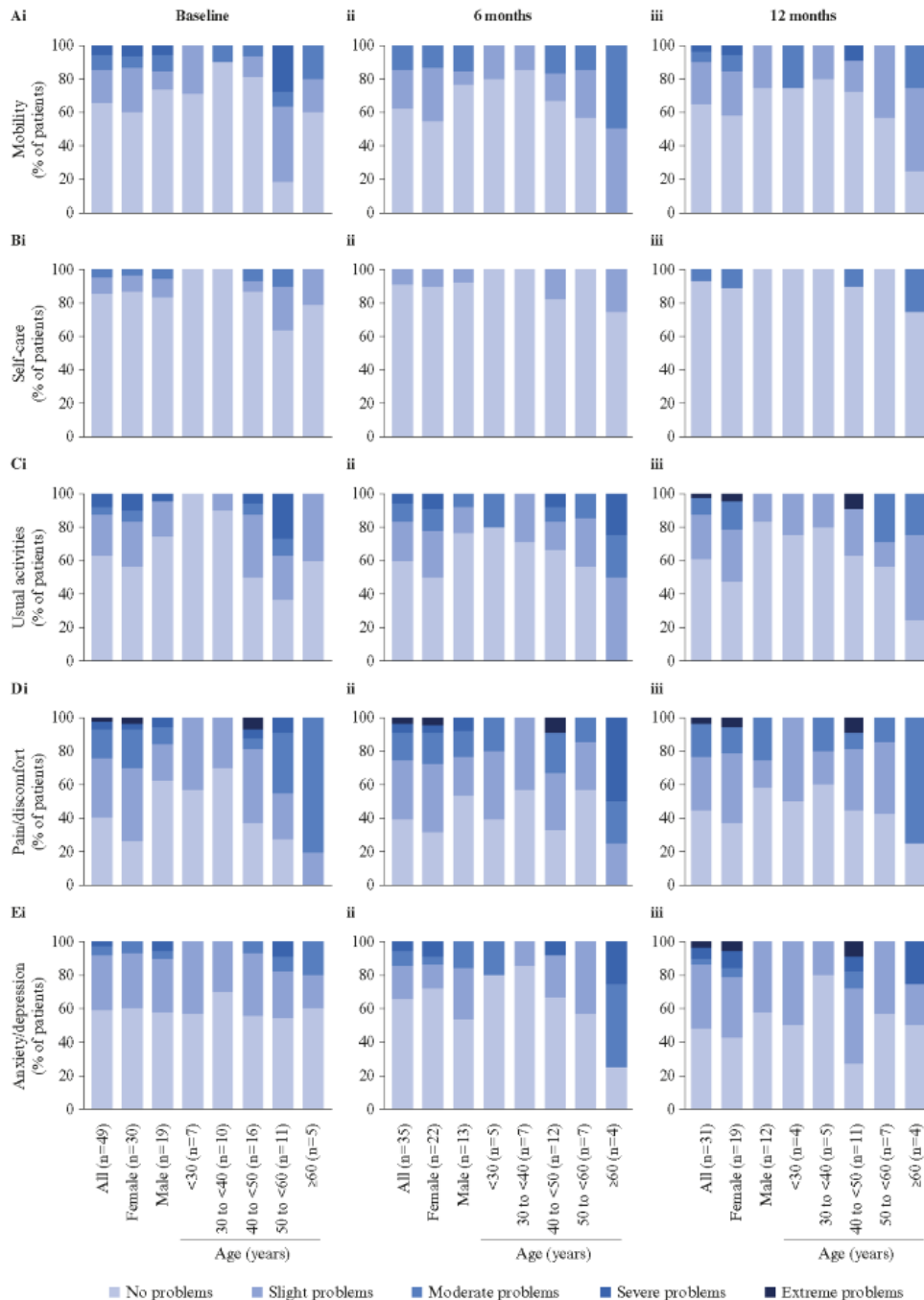
The median TSQM (version II) global satisfaction domain score remained stable during follow-up, showed no consistent difference between men and women, and was lowest in the youngest age group during follow-up (Table 3). Median domain scores for effectiveness and side effects showed no consistent differences between the age and gender subgroups. The median domain score for convenience was lower in women than in men and was lowest in the youngest age group (Table 3).

Most of the patients were satisfied or *neither satisfied nor dissatisfied* with the BETAPLUS PSDMP (Multimedia Appendix 2). Most patients were satisfied or *neither satisfied nor dissatisfied* with the BETACONNECT (Multimedia

Appendix 3) and satisfied or very satisfied with the app (Multimedia Appendix 4). In analyses stratified by gender and age groups, the sample sizes in the respective strata were very small and did not allow for any further reliable conclusions (Multimedia Appendices 2-4).

The baseline EuroQol VAS score was higher in persistent and adherent patients than in nonpersistent and nonadherent patients (Multimedia Appendix 5). Logistic regression showed a significant association of the baseline EuroQol VAS score with persistence and adherence at 12 months but not at 6 months (Multimedia Appendix 6).

Table 3. Treatment Satisfaction Questionnaire for Medication (version II) domain scores.

Characteristics	Baseline		Month 6		Month 12	
	n (%)	Median (IQR)	n (%)	Median (IQR)	n (%)	Median (IQR)
Effectiveness						
Total (N=62)	49 (79)	83.3 (66.7-91.7)	35 (56)	83.3 (58.3-91.7)	31 (50)	83.3 (58.3-100.0)
Gender						
Female (n=41)	30 (73)	83.3 (66.7-91.7)	22 (54)	83.3 (66.7-91.7)	19 (46)	83.3 (58.3-100.0)
Male (n=21)	19 (90)	75.0 (58.3-88.3)	13 (62)	83.3 (41.7-100.0)	12 (57)	83.3 (70.8-95.8)
Age (years)						
<30 (n=9)	7 (78)	66.7 (66.7-100.0)	5 (56)	83.3 (75.0-91.7)	4 (44)	66.7 (41.7-83.3)
30 to <40 (n=15)	10 (67)	83.3 (83.3-83.3)	7 (47)	83.3 (75.0-91.7)	5 (33)	91.7 (83.3-91.7)
40 to <50 (n=19)	16 (84)	83.3 (79.2-91.7)	12 (63)	83.3 (50.0-91.7)	11 (58)	83.3 (66.7-100.0)
50 to <60 (n=13)	11 (85)	75.0 (50.0-91.7)	7 (54)	75.0 (25.0-100.0)	7 (54)	91.7 (25.0-100.0)
≥60 (n=6)	5 (83)	66.7 (58.3-66.7)	4 (67)	66.7 (54.2-79.2)	4 (67)	91.7 (70.8-100.0)
Side effects						
Total (N=62)	29 (47)	75.0 (58.3-83.3)	18 (29)	75.0 (58.3-87.5)	15 (24)	75.0 (75.0-91.7)
Gender						
Female (n=41)	16 (39)	70.8 (45.8-83.3)	10 (24)	75.0 (50.0-91.7)	7 (17)	75.0 (75.0-91.7)
Male (n=21)	13 (62)	75.0 (66.7-87.5)	8 (38)	75.0 (60.4-79.2)	8 (38)	75.0 (70.8-87.5)
Age (years)						
<30 (n=9)	3 (33)	83.3 (41.7-83.3)	2 (22)	75.0 (75.0-75.0)	0 (0)	N/A[a]
30 to <40 (n=15)	7 (47)	75.0 (50.0-83.3)	4 (27)	54.2 (37.5-75.0)	2 (13)	75.0 (75.0-75.0)
40 to <50 (n=19)	8 (42)	70.8 (62.5-87.5)	6 (32)	66.7 (58.3-83.3)	5 (26)	83.3 (75.0-91.7)
50 to <60 (n=13)	7 (54)	83.3 (62.5-100.0)	4 (31)	81.3 (68.8-89.6)	5 (38)	75.0 (75.0-83.3)
≥60 (n=6)	4 (67)	62.5 (45.8-79.2)	2 (33)	87.5 (75.0-100.0)	3 (50)	75.0 (58.3-91.7)
Convenience						
Total (N=62)	49 (79)	72.2 (61.1-83.3)	35 (56)	77.8 (66.7-88.9)	31 (50)	72.2 (61.1-88.9)
Gender						
Female (n=41)	30 (73)	69.4 (61.1-77.8)	22 (54)	75.0 (55.6-83.3)	19 (46)	72.2 (50.0-88.9)
Male (n=21)	19 (90)	77.8 (66.7-88.9)	13 (62)	83.3 (77.8-88.9)	12 (57)	77.8 (66.7-86.1)
Age (years)						
<30 (n=9)	7 (78)	66.7 (55.6-72.2)	5 (56)	55.6 (22.2-66.7)	4 (44)	47.2 (36.1-55.6)
30 to <40 (n=15)	10 (67)	66.7 (61.1-77.8)	7 (47)	77.8 (66.7-77.8)	5 (33)	72.2 (66.7-83.3)
40 to <50 (n=19)	16 (84)	72.2 (66.7-83.3)	12 (63)	83.3 (69.4-88.9)	11 (58)	77.8 (50.0-88.9)
50 to <60 (n=13)	11 (85)	88.9 (72.2-94.4)	7 (54)	88.9 (77.8-94.4)	7 (54)	88.9 (77.8-94.4)
≥60 (n=6)	5 (83)	72.2 (61.1-72.2)	4 (67)	75.0 (58.3-91.7)	4 (67)	72.2 (63.9-80.6)
Global satisfaction						
Total (N=62)	49 (79)	83.3 (66.7-91.7)	35 (56)	83.3 (66.7-91.7)	31 (50)	83.3 (66.7-91.7)
Gender						
Female (n=41)	30 (73)	83.3 (66.7-91.7)	22 (54)	83.3 (66.7-91.7)	19 (46)	75.0 (58.3-91.7)
Male (n=21)	19 (90)	83.3 (66.7-91.7)	13 (62)	75.0 (66.7-91.7)	12 (57)	83.3 (70.8-91.7)
Age (years)						
<30 (n=9)	7 (78)	75.0 (66.7-83.3)	5 (56)	66.7 (58.3-91.7)	4 (44)	58.3 (54.2-66.7)
30 to <40 (n=15)	10 (67)	83.3 (66.7-83.3)	7 (47)	75.0 (66.7-83.3)	5 (33)	83.3 (75.0-91.7)

XSL·FO
RenderX

Characteristics	Baseline		Month 6		Month 12	
	n (%)	Median (IQR)	n (%)	Median (IQR)	n (%)	Median (IQR)
40 to <50 (n=19)	16 (84)	79.2 (66.7-87.5)	12 (63)	79.2 (66.7-91.7)	11 (58)	83.3 (66.7-83.3)
50 to <60 (n=13)	11 (85)	83.3 (66.7-100.0)	7 (54)	83.3 (66.7-100.0)	7 (54)	91.7 (83.3-100.0)
≥60 (n=6)	5 (83)	75.0 (66.7-75.0)	4 (67)	79.2 (66.7-91.7)	4 (67)	79.2 (70.8-83.3)

[a]N/A: not applicable.

Baseline TSQM (version II) domain scores for effectiveness, side effects, convenience, and global satisfaction were higher in persistent and adherent patients than in nonpersistent and nonadherent patients (Multimedia Appendix 7). Logistic regression showed a significant association between the baseline side effects score and persistence at 12 months, between the baseline convenience score and adherence at 6 and 12 months, and between the baseline global satisfaction score and adherence at 6 and 12 months (Multimedia Appendix 6).

The proportion of patients who were satisfied or very satisfied with the PSDMP at baseline was higher in persistent patients than in nonpersistent patients, whereas it was similar in adherent and nonadherent patients (Multimedia Appendix 8). Regarding the BETACONNECT and the app, the proportion of patients who were satisfied or very satisfied at baseline was higher in persistent and adherent patients than in nonpersistent and nonadherent patients (Multimedia Appendix 8). Logistic regression analyses did not suggest significant associations of persistence and adherence with satisfaction with the PSDMP, BETACONNECT, or app (Multimedia Appendix 6).

PEAK Use

The proportion of patients using PEAK at least once was higher among men (14/21, 67%) than among women (20/41, 49%) and was lowest among those aged ≥60 years (1/6, 17% compared with 6/9, 67%; 9/15, 60%; 10/19, 53%; and 8/13, 62% among those aged <30; 30 to <40; 40 to <50; and 50 to <60 years, respectively).

In an analysis of PEAK use over 3-month intervals, the number of active days (days in which PEAK was used at least once) and the average number of games played per week varied widely between patients in each interval (Table 4).

The 10 most popular games included four games in the language category (Word Hunt, Babble Bots, Word Fresh, and Grow), two in the focus category (Must Sort and Objectifind), two in the problem-solving category (Pixel Logic and Low Pop), one in the memory category (Perilous Path), and one in the mental agility category (Turtle Traffic). Of the 10 most popular games, Pixel Logic was played for the longest cumulative time during each 3-month interval (median 37, IQR 3-107 minutes at months 1-3; median 6, IQR 0-99 minutes at months 4-6; median 11, IQR 0-55 minutes at months 7-9; and median 25, IQR 2-59 minutes at months 10-12).

Table 4. PEAK use over 3-month intervals.

Characteristics	Values (N=62)		
	n (%)	Median (IQR)	Range
Number of days active per patient[a]			
Data not assigned to any time interval	12 (19)	5 (2-37)	1-83
Months 1-3	32 (52)	23 (8-49)	1-91
Months 4-6	18 (29)	4 (1-68)	1-91
Months 7-9	14 (23)	14 (1-60)	1-81
Months 10-12	14 (23)	10 (2-27)	2-89
Number of games played per patient per week[b]			
Data not assigned to any time interval	12 (19)	0 (0-8)	0-33
Months 1-3	32 (52)	5 (1-12)	0-158
Months 4-6	18 (29)	0 (0-11)	0-218
Months 7-9	14 (23)	2 (0-10)	0-259
Months 10-12	14 (23)	4 (0-7)	0-116

[a]An active day was defined as a day on which PEAK was used at least once.

[b]Only the 10 most frequently played games (all PEAK users) were counted. A single game could be played multiple times per day.

Discussion

Principal Findings

This study builds on the results of a previous pilot study [5], performing follow-up for 1 year, and incorporating completion of PRO instruments.

The proportion of patients participating was 4.44% (79/1778) of registered accounts for the app. More than three-quarters (48/62, 77%) of the participants had ≥12 months of injection data for prospective analysis. Compliance over the 1-year prospective observational period was high (median 98.9%, IQR 94.3%-100%), with similar percentages among men and women. Persistence and adherence decreased from 85% (53/62) and 74% (46/62), respectively, at 6 months to 76% (47/62) and 65% (40/62), respectively, at 12 months, and were higher in men than in women in the prospective analysis.

The PRO questionnaires were answered by 79% (49/62) of the participants at baseline and 50% (31/62) of the participants at month 12. An immediate reward for participation rather than compensation at the end of the study may have increased the completion rate of the PRO questionnaires [11,12]. Women reported more severe problems than men in some EQ-5D-5L dimensions (mobility, usual activities, and pain/discomfort) and had lower median domain scores than men for convenience in the TSQM (version II). HRQoL showed no consistent changes across age groups. Convenience and global satisfaction (assessed using the TSQM [version II]) were lowest in the youngest age group. Persistence and adherence at 12 months were associated with baseline HRQoL (EuroQol VAS) and satisfaction with treatment (scores in specific domains of the TSQM [version II]) but not satisfaction with support.

PEAK was used at least once by 67% (14/21) of the male participants and by 49% (20/41) of the female participants. The frequency of PEAK use varied widely among patients.

Comparison With Previous Work

This study used the same digital approach as the previous pilot study [5] but recorded data over a longer period (1 year prospectively and ≤1 year retrospectively compared with 6 months in total in the pilot study) and allowed comparison of medication-taking behavior between retrospective and prospective study periods. In addition, this study recorded PROs and tested the acceptance of an optional cognitive training tool (ie, it involved active engagement by the participants), whereas the pilot study used only basic registration data and routinely recorded injection data.

The pattern of decreasing persistence and adherence with continuing high compliance over the 1-year follow-up period appeared in both the prospective and retrospective analyses. These findings confirm previous reports that an increasing number of patients stop their treatment over time [13-16], with substantial attrition occurring early in the treatment course, within 1 year [15,16]. In contrast, the high compliance at both 6 and 12 months suggests that patients remaining on treatment take their medication very reliably. The similarity in persistence, compliance, and adherence between the prospective and retrospective analyses argues against an observer effect prompting patients to act differently once they were aware that they were participating in a study (ie, after providing consent).

Persistence and adherence in the prospective observational period (53/62, 85% and 46/62, 74%, respectively, at 6 months) were lower than those reported in the pilot study (90/94, 96% and 84/94, 89%, respectively) [5]. This may be explained by patient motivation and stamina, which may be higher in a 6-month study than in a 1-year study.

Men appeared more persistent and more adherent than women in both prospective and retrospective analyses. We cannot say whether this is a valid finding or a result of confounding; for example, by men being more technophile than women. In this context, more women may have chosen to stop documenting injections in the app over time, which by definition would have resulted in their classification as nonpersistent and nonadherent thereafter.

Persistence and adherence showed little difference between men and women in the pilot study [5] and in an observational study of patients receiving interferon β-1b via the BETACONNECT autoinjector in Germany (the BETAPREDICT study) [13]. However, several large studies of patients with MS have shown greater adherence in men than in women [17-19], consistent with our current findings. A US-based administrative claims database study of 648 patients with MS showed that those who were adherent to disease-modifying therapy over a study period of ≥24 months were more likely to be men than in nonadherent patients (173/448, 38.6% vs 52/200, 26%; *P*=.002) [17]. Another US-based study used enrollment and claims data from an upper Midwest health plan between 2011 and 2013 and showed that female patients were 5.5% less likely to be adherent than male patients [18]. A third US-based claims database study of 8382 patients with MS followed for 12 months found that male patients had an increased likelihood of adherence to disease-modifying therapy (odds ratio 1.2, 95% CI 1.085-1.335; *P*<.001) [19].

We found the lowest rates of adherence and global treatment satisfaction in the youngest age group (<30 years) in the prospective analysis. In the pilot study, those aged 30 to 39 years had the lowest rate of adherence [5]. Although age thresholds vary across studies, these results are in general agreement with other studies reporting higher rates of adherence in older patients [17-19]. A US-based administrative claims database study showed that patients with MS who were adherent to disease-modifying therapy were on average older than those who were nonadherent (mean 43.5, SD 8.0 years vs mean 41.8, SD 8.1 years; *P*=.02) [17]. A US-based study using health plan data showed that patients aged >45 years were 13.7% to 18.6% more likely to be adherent than younger patients [18]. A third US-based claims database study found that age groups older than 18-34 years had an increased likelihood of adherence to disease-modifying therapy (odds ratios 1.220-1.331; *P*<.001 to *P*=.001) [19].

Women reported problems with mobility, usual activities, and pain/discomfort more frequently than men in our analysis. A previous study of 144 patients (99 women and 45 men) with MS in Germany also found that women reported problems with usual activities more frequently than men, although the

association was not significant in multivariate analysis [20]. In contrast, a large UK-based study of 4516 patients with MS (including 3198 women and 1301 men) showed that men reported problems with mobility, self-care, and usual activities more frequently than women [21]. The reasons for the inconsistency between studies are unclear but could include differences in study design and location.

There was a steep attrition of participants' use of the PEAK app after 3 months, as also seen for other apps [22]. Free access to the professional version of the brain training tool did not provide enough incentive for participants to continue using it over the entire study period. Hence, voluntary and free access to PEAK does not appear to be an effective approach to provide constant brain training over a longer period. Peer-to-peer interactions (eg, via social media challenges, multiplayer games, or guided training programs) may have a positive impact on long-term engagement with gamified apps [23-25].

Limitations

The generalizability of the results is limited by the small sample size and the possibility of selection bias. The number of patients was lower than in the pilot study [5], and it is possible that mainly technophile patients agreed to participate. We were unable to compare app users who participated in the study with those who did not, because demographic and clinical data from the latter group were not available.

Although we provided incentives (vouchers) for the completion of each questionnaire after the end of the observation period, the number of participants still decreased over the 1-year period. More immediate ways to incentivize patients directly after completing the questionnaires may be warranted to encourage continued active study participation.

Only patients using interferon β-1b were eligible to participate in this study. However, the aim of our study was to describe medication-taking behavior and collect PROs among patients treated with interferon β-1b and not to compare their behavior with that of other patient groups. In the future, open platforms that do not restrict participation to users of certain drugs will be more versatile in terms of patient pool, study design, and indication than drug-specific platforms. IDS and Bayer have codeveloped a data capture tool, the my ePRO app, which can be used for stand-alone studies or in association with randomized controlled trials and observational studies. The usability of the my ePRO app was assessed in the DePRO study [26]. The DePRO study showed that digitally authenticating eligible patients, with participants registering by scanning the 2D matrix code on the outer packaging of their prescribed medication (a standard feature of prescription medications in the European Union) is a feasible approach for a digital study (manuscript in preparation).

We were unable to perform source data verification or clarify incomplete data from individual patients because the analysis results were anonymized. We were also unable to assess whether offering access to PEAK had any influence on adherence or commitment to the study because of the lack of a control group. Other limitations are similar to those described for the pilot study [5], including the use of nonvalidated technology to obtain data on medication-taking behavior, missing data (which have not been replaced), and the inability to distinguish between patients discontinuing treatment and those simply ceasing to document their injections.

Conclusions

This study showed high compliance and decreasing persistence and adherence over 1 year and demonstrated the feasibility of including remotely completed electronic PRO instruments in digital observational studies. The feasibility and influence of gamification in this setting remain unclear.

Conflicts of Interest

VL has served as an adviser or speaker or received research grants from Antisense, Allergan, Bayer, Biogen, Genzyme, Novartis, Roche, and TEVA. CM and MS are full-time employees of Bayer Vital GmbH (Leverkusen, Germany). KB-G is an employee of the Institut Dr. Schauerte (Munich, Germany). Bayer selected the Institut Dr. Schauerte for the statistical analysis of the study.

Multimedia Appendix 1
Service questionnaire.
[DOC File , 46 KB - formative_v6i3e31972_app1.doc]

Multimedia Appendix 2
Responses to service questionnaire part 1: "Are you satisfied with the BETAPLUS patient support program?".
[DOC File , 137 KB - formative_v6i3e31972_app2.doc]

Multimedia Appendix 3
Responses to service questionnaire part 2: "Are you satisfied with the BETACONNECT autoinjector?".

[DOC File , 140 KB - formative_v6i3e31972_app3.doc]

Multimedia Appendix 4
Responses to service questionnaire part 3: "Are you satisfied with the myBETAapp?".
[DOC File , 124 KB - formative_v6i3e31972_app4.doc]

Multimedia Appendix 5
Baseline EuroQol 5-Dimension, 5-Level questionnaire (EQ-5D-5L) visual analogue scale stratified by persistence and adherence at 6 and 12 months.
[DOC File , 78 KB - formative_v6i3e31972_app5.doc]

Multimedia Appendix 6
Association of persistence and adherence at 6 and 12 months with baseline health-related quality of life and satisfaction with treatment and support.
[DOC File , 134 KB - formative_v6i3e31972_app6.doc]

Multimedia Appendix 7
Baseline Treatment Satisfaction Questionnaire for Medication (version II) domain scores stratified by persistence and adherence at 6 and 12 months.
[DOC File , 164 KB - formative_v6i3e31972_app7.doc]

Multimedia Appendix 8
Baseline satisfaction with the Patient Support and Disease Management Program, BETACONNECT, and app stratified by persistence and adherence at 6 and 12 months.
[DOC File , 99 KB - formative_v6i3e31972_app8.doc]

References

1. Inan OT, Tenaerts P, Prindiville SA, Reynolds HR, Dizon DS, Cooper-Arnold K, et al. Digitizing clinical trials. NPJ Digit Med 2020;3:101 [FREE Full text] [doi: 10.1038/s41746-020-0302-y] [Medline: 32821856]
2. Bower P, Brueton V, Gamble C, Treweek S, Smith CT, Young B, et al. Interventions to improve recruitment and retention in clinical trials: a survey and workshop to assess current practice and future priorities. Trials 2014;15:399 [FREE Full text] [doi: 10.1186/1745-6215-15-399] [Medline: 25322807]
3. Moore TJ, Heyward J, Anderson G, Alexander GC. Variation in the estimated costs of pivotal clinical benefit trials supporting the US approval of new therapeutic agents, 2015-2017: a cross-sectional study. BMJ Open 2020 Jun 11;10(6):e038863 [FREE Full text] [doi: 10.1136/bmjopen-2020-038863] [Medline: 32532786]
4. Unger JM, Cook E, Tai E, Bleyer A. The role of clinical trial participation in cancer research: barriers, evidence, and strategies. Am Soc Clin Oncol Educ Book 2016;35:185-198. [doi: 10.14694/EDBK_156686] [Medline: 27249699]
5. Limmroth V, Hechenbichler K, Müller C, Schürks M. Assessment of medication adherence using a medical app among patients with multiple sclerosis treated with interferon beta-1b: Pilot Digital Observational Study (PROmyBETAapp). J Med Internet Res 2019 Jul 29;21(7):e14373 [FREE Full text] [doi: 10.2196/14373] [Medline: 31359863]
6. PROmyBETAappGame: a study to learn more about the medication usage and patient reported outcomes via the myBETAapp and to find out more about the usage of game principles and game design elements (Gamification) in medical care of patients with multiple sclerosis treated with Betaferon. ClinicalTrials. 2019. URL: https://clinicaltrials.gov/ct2/show/NCT03808142 [accessed 2022-02-21]
7. Chiaravalloti ND, DeLuca J. Cognitive impairment in multiple sclerosis. Lancet Neurol 2008 Dec;7(12):1139-1151. [doi: 10.1016/S1474-4422(08)70259-X] [Medline: 19007738]
8. Atkinson MJ, Kumar R, Cappelleri JC, Hass SL. Hierarchical construct validity of the treatment satisfaction questionnaire for medication (TSQM version II) among outpatient pharmacy consumers. Value Health 2005 Dec;8 Suppl 1:9-24 [FREE Full text] [doi: 10.1111/j.1524-4733.2005.00066.x] [Medline: 16336491]
9. Atkinson MJ, Sinha A, Hass SL, Colman SS, Kumar RN, Brod M, et al. Validation of a general measure of treatment satisfaction, the Treatment Satisfaction Questionnaire for Medication (TSQM), using a national panel study of chronic disease. Health Qual Life Outcomes 2004 Feb 26;2:12 [FREE Full text] [doi: 10.1186/1477-7525-2-12] [Medline: 14987333]
10. Washington F, Langdon D. Factors affecting adherence to disease-modifying therapies in multiple sclerosis: systematic review. J Neurol 2021;epub ahead of print. [doi: 10.1007/s00415-021-10850-w] [Medline: 34676448]
11. Reynolds B, Schiffbauer R. Delay of gratification and delay discounting: a unifying feedback model of delay-related impulsive behavior. Psychol Rec 2005;55(3):439-460. [doi: 10.1007/bf03395520]

XSL•FO
RenderX

12. Yu R. Regional white matter volumes correlate with delay discounting. PLoS One 2012;7(2):e32595 [FREE Full text] [doi: 10.1371/journal.pone.0032595] [Medline: 22393420]

13. Köhler W, Bayer-Gersmann K, Neußer T, Schürks M, Ziemssen T. Predictors of adherence among patients with multiple sclerosis using the BETACONNECT Autoinjector: a prospective observational cohort study. Front Neurol 2021;12:643126 [FREE Full text] [doi: 10.3389/fneur.2021.643126] [Medline: 33716945]

14. Kleiter I, Lang M, Jeske J, Norenberg C, Stollfuß B, Schürks M. Adherence, satisfaction and functional health status among patients with multiple sclerosis using the BETACONNECT® autoinjector: a prospective observational cohort study. BMC Neurol 2017 Sep 06;17(1):174 [FREE Full text] [doi: 10.1186/s12883-017-0953-8] [Medline: 28877664]

15. Hansen K, Schüssel K, Kieble M, Werning J, Schulz M, Friis R, et al. Adherence to disease modifying drugs among patients with multiple sclerosis in Germany: a retrospective cohort study. PLoS One 2015;10(7):e0133279 [FREE Full text] [doi: 10.1371/journal.pone.0133279] [Medline: 26214805]

16. Halpern R, Agarwal S, Dembek C, Borton L, Lopez-Bresnahan M. Comparison of adherence and persistence among multiple sclerosis patients treated with disease-modifying therapies: a retrospective administrative claims analysis. Patient Prefer Adher 2011 Jan 20;5:73-84 [FREE Full text] [doi: 10.2147/PPA.S15702] [Medline: 21423591]

17. Ivanova JI, Bergman RE, Birnbaum HG, Phillips AL, Stewart M, Meletiche DM. Impact of medication adherence to disease-modifying drugs on severe relapse, and direct and indirect costs among employees with multiple sclerosis in the US. J Med Econ 2012;15(3):601-609. [doi: 10.3111/13696998.2012.667027] [Medline: 22376190]

18. Higuera L, Carlin CS, Anderson S. Adherence to disease-modifying therapies for multiple sclerosis. J Manag Care Spec Pharm 2016 Dec;22(12):1394-1401 [FREE Full text] [doi: 10.18553/jmcp.2016.22.12.1394] [Medline: 27882830]

19. Munsell M, Frean M, Menzin J, Phillips AL. An evaluation of adherence in patients with multiple sclerosis newly initiating treatment with a self-injectable or an oral disease-modifying drug. Patient Prefer Adher 2017;11:55-62 [FREE Full text] [doi: 10.2147/PPA.S118107] [Medline: 28115831]

20. Reese JP, Wienemann G, John A, Linnemann A, Balzer-Geldsetzer M, Mueller UO, et al. Preference-based Health status in a German outpatient cohort with multiple sclerosis. Health Qual Life Outcomes 2013 Oct 03;11:162 [FREE Full text] [doi: 10.1186/1477-7525-11-162] [Medline: 24089999]

21. Jones KH, Ford DV, Jones PA, John A, Middleton RM, Lockhart-Jones H, et al. How people with multiple sclerosis rate their quality of life: an EQ-5D survey via the UK MS register. PLoS One 2013;8(6):e65640 [FREE Full text] [doi: 10.1371/journal.pone.0065640] [Medline: 23776516]

22. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

23. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support and social media. Epidemiol Psychiatr Sci 2016 Apr;25(2):113-122 [FREE Full text] [doi: 10.1017/S2045796015001067] [Medline: 26744309]

24. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Dec;71:31-48. [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]

25. Santos LH, Okamoto K, Funghetto SS, Cavalli AS, Hiragi S, Yamamoto G, et al. Effects of social interaction mechanics in pervasive games on the physical activity levels of older adults: quasi-experimental study. JMIR Serious Games 2019 Jul 22;7(3):e13962 [FREE Full text] [doi: 10.2196/13962] [Medline: 31333202]

26. Mueller C, Schauerte I, Martin S. Evaluation of self-care activities and quality of life in patients with type 2 diabetes mellitus treated with metformin using the 2D matrix code of outer drug packages as patient identifier: protocol for the DePRO proof-of-concept observational study. JMIR Res Protoc 2021 Jan 11;10(1):e21727 [FREE Full text] [doi: 10.2196/21727] [Medline: 33427685]

Abbreviations

HRQoL: health-related quality of life
IDS: Institut Dr. Schauerte
MS: multiple sclerosis
PRO: patient-reported outcome
PSDMP: Patient Support and Disease Management Program
TSQM: Treatment Satisfaction Questionnaire for Medication
VAS: visual analogue scale

XSL•FO

RenderX

Machine-Aided Self-diagnostic Prediction Models for Polycystic Ovary Syndrome: Observational Study

Angela Zigarelli[1], MSc; Ziyang Jia[1], MSc; Hyunsun Lee[1], PhD

Department of Mathematics and Statistics, University of Massachusetts Amherst, Newton, MA, United States

Corresponding Author:
Hyunsun Lee, PhD
Department of Mathematics and Statistics
University of Massachusetts Amherst
100 Carlson Ave
Newton, MA, 02459
United States
Phone: 1 6172431886
Email: hyunsunlee@umass.edu

Abstract

Background: Artificial intelligence and digital health care have substantially advanced to improve and enhance medical diagnosis and treatment during the prolonged period of the COVID-19 global pandemic. In this study, we discuss the development of prediction models for the self-diagnosis of polycystic ovary syndrome (PCOS) using machine learning techniques.

Objective: We aim to develop self-diagnostic prediction models for PCOS in potential patients and clinical providers. For potential patients, the prediction is based only on noninvasive measures such as anthropomorphic measures, symptoms, age, and other lifestyle factors so that the proposed prediction tool can be conveniently used without any laboratory or ultrasound test results. For clinical providers who can access patients' medical test results, prediction models using all predictor variables can be adopted to help health providers diagnose patients with PCOS. We compare both prediction models using various error metrics. We call the former model the patient model and the latter, the provider model throughout this paper.

Methods: In this retrospective study, a publicly available data set of 541 women's health information collected from 10 different hospitals in Kerala, India, including PCOS status, was acquired and used for analysis. We adopted the CatBoost method for classification, *K*-fold cross-validation for estimating the performance of models, and SHAP (Shapley Additive Explanations) values to explain the importance of each variable. In our subgroup study, we used *k*-means clustering and Principal Component Analysis to split the data set into 2 distinct BMI subgroups and compared the prediction results as well as the feature importance between the 2 subgroups.

Results: We achieved 81% to 82.5% prediction accuracy of PCOS status without any invasive measures in the patient models and achieved 87.5% to 90.1% prediction accuracy using both noninvasive and invasive predictor variables in the provider models. Among noninvasive measures, variables including acanthosis nigricans, acne, hirsutism, irregular menstrual cycle, length of menstrual cycle, weight gain, fast food consumption, and age were more important in the models. In medical test results, the numbers of follicles in the right and left ovaries and anti-Müllerian hormone were ranked highly in feature importance. We also reported more detailed results in a subgroup study.

Conclusions: The proposed prediction models are ultimately expected to serve as a convenient digital platform with which users can acquire pre- or self-diagnosis and counsel for the risk of PCOS, with or without obtaining medical test results. It will enable women to conveniently access the platform at home without delay before they seek further medical care. Clinical providers can also use the proposed prediction tool to help diagnose PCOS in women.

KEYWORDS

XSL·FO

RenderX

Introduction

Background

There has been substantial advancement in artificial intelligence and machine learning technologies in health care owing to the prolonged COVID-19 pandemic, resulting in improvements and enhancements in medical diagnosis and treatment that were previously impossible or unavailable [1]. Telehealth using remote technologies between medical providers and patients is another emerging trend during the pandemic, and many conditions are diagnosed and managed through telehealth, including polycystic ovary syndrome (PCOS) [2,3]. Diagnosis of PCOS with telehealth is based on a few symptoms such as irregular menstrual cycles, hirsutism, skin problems, and other symptoms caused by an imbalance of androgen hormones [3]. Through our proposed study, integrating the current trends addressed above, we provide a more systematic self-diagnostic tool that allows users to conveniently access and learn the predicted diagnosis result of PCOS without delay before they seek further medical care.

PCOS is the most common endocrine disorder among women of reproductive age, possibly causing infertility. The prevalence of PCOS ranges from 6% to 20%, depending on the population and the diagnostic criteria reported in previous studies [4-6]. Although the cause of this syndrome is not clearly known, there is increasing evidence that PCOS is a complex multigenic disorder with strong epigenetic and environmental influences [4]. According to the Centers for Disease Control and Prevention, overweight women with PCOS may develop serious health problems such as diabetes, gestational diabetes, heart disease, high blood pressure (BP), sleep apnea, and stroke. PCOS is also known to be linked to anxiety and depression [7]. It is important to note that not all women with PCOS experience the same combination or severity of symptoms, which makes early detection more challenging [8].

Although several criteria for PCOS diagnosis have been proposed, the Rotterdam criteria were accepted by the National Institutes of Health, and they have been most commonly adopted for the diagnosis of PCOS [9]. On the basis of the Rotterdam criteria, the diagnosis is made if at least two out of three of the following criteria are met: ovulatory dysfunction (oligo-ovulation or anovulation), higher levels of androgens in the blood, and polycystic ovaries appearing on ultrasound. However, the Rotterdam criteria have been controversial in many studies [9-11].

There are two types of PCOS: lean and obese, each with different biochemical, hormonal, and metabolic profiles [12]. Toosy et al [8] noted that a smaller proportion of women with lean PCOS had a normal or low BMI (≤ 25 kg/m^2) and may or may not have symptoms, which makes diagnosis more challenging. In other studies, it has been noted that PCOS is closely associated with obesity and is more prevalent among overweight or obese women than in the general population of women of reproductive age [13-15]. In our proposed study, we incorporate the ideas of lean and obese PCOS types in earlier studies and group the data set into 2 subgroups based on anthropomorphic measures.

Objectives

Machine learning and deep learning techniques have been widely used to analyze health data and improve diagnostic accuracy and precision, disease treatment, and prevention [16,17]. In particular, feature selection, clustering algorithms, and classification have often been adopted for subgroup studies [18,19]. The goal of this proposed study is to develop a machine-aided self-diagnostic tool that predicts the diagnosis of PCOS with and without any invasive measures, using Principal Component Analysis (PCA), *k*-means clustering algorithm, and CatBoost classifier. The CatBoost method is one of the newer gradient boosting decision tree models, and it was recently used in diabetes prediction in the study by Kumar et al [20]. Our development ultimately enables users, either potential patients or clinical providers, to conveniently access this pre- or self-diagnostic digital platform for PCOS from anywhere. This work is well aligned with emerging artificial intelligence and digital health care [21].

The remainder of this paper is organized as follows. In the *Methods* section, we discuss data preparation, provide an overview of statistical analysis, and explain each machine learning technique used in our analysis. In the *Results* section, we report our findings in a subgroup study and evaluate the performance of our proposed prediction models. In the *Discussion* section, the major results are highlighted, and we conclude this paper.

Methods

Subjects and Data Preparation

In this retrospective study, we obtained and analyzed a publicly available data set that was collected from 10 different hospitals across Kerala, India [22]. After data cleaning and removing 15 partly missing or high-leverage data points, the study cohort consisted of 526 women aged between 20 and 48 years, of which 170 (32.3%) were diagnosed with PCOS and 356 (67.7%) were not diagnosed with PCOS. The data set included other health information for each subject, such as anthropomorphic attributes, symptoms, laboratory and ultrasound test results, age, blood type, marital status, pregnancy, abortion history, fast food consumption, and exercise. For later use, we classified and named these variables as *anthropomorphic* variables, *symptom* variables, *test result* variables, and *given* variables in this paper.

The *anthropomorphic* variables include six variables: BMI, height, hip circumference, waist circumference, waist-to-hip ratio, and body weight. The *symptom* variables are self-observable variables. The seven symptom variables are acanthosis nigricans (skin darkening in body folds and creases), acne, hair loss, hirsutism, irregular menstrual cycle, length of menstrual cycle, and weight gain. The *test result* variables are based on blood work and ultrasound or any other medical examination, and they include anti-Müllerian hormone (AMH), the number of antral follicles in the left ovary, the number of antral follicles in the right ovary, average follicle size in the left ovary, average follicle size in the right ovary, diastolic BP, endometrium thickness, follicle-stimulating hormone, follicle-stimulating hormone to luteinizing hormone ratio, glycated hemoglobin levels, human chorionic gonadotropin I,

human chorionic gonadotropin II, luteinizing hormone, progesterone, prolactin, pulse rate, random glucose test, respiratory rate, systolic BP, thyrotropin, and vitamin D3. The remaining variables in the data set, other than the status of PCOS and the variables mentioned above, are defined as *given* variables which are age, blood type, years of marriage, abortion history, fast food consumption, pregnancy status, and regular exercise.

Overview of Statistical Analysis

We first examined the difference in attributes between the PCOS-positive group and the PCOS-negative group as a preliminary investigation. In our analysis, we provided two types of prediction models, with and without invasive health measures. In a model, we used only noninvasive variables, that is, anthropomorphic, symptom and given variables, and called this model the *patient* model. In the other model, we used all variables including noninvasive and invasive variables, and called this model the *provider* model. The purpose of proposing both prediction models was to accommodate various users who may or may not have had access to medical test results.

Another part of the statistical analysis was a subgroup study. There have been several subgroup studies of PCOS diagnosis based on different combinations of the 3 Rotterdam criteria [23,24]. Kar et al [24] characterized and classified phenotypes of PCOS in a large cohort of women into subgroups and compared the data of various metabolic complications of these phenotypes. In our subgroup study, we divided 526 women into 2 subgroups based on anthropomorphic attributes. The motivation for this idea originated from predicting lean PCOS and obese PCOS in a more detailed manner. In the remainder of this section, we explain each machine learning technique and the application of these techniques in our analysis.

PCA Method

The PCA is a dimension reduction method that is often used when there are highly correlated variables in the data set. It increases interpretability to a certain extent and minimizes data loss by combining correlated variables together to create a new set of uncorrelated yet more representative variables [25,26]. We use PCA to perform feature extraction, where several highly correlated anthropomorphic variables are linearly combined to create new axes, that is, the principal components. For example, body weight and BMI are highly correlated and are consequently highly related to the same principal component. The number of principal components for the analysis can be determined by examining the proportion of variance explained and the elbow rule.

K-Means Clustering

We adopted *k*-means clustering, which is an unsupervised machine learning algorithm that makes inferences using only input variables without referring to known outcomes. It is often used to understand the latent structure within a data set by aggregating data points based on certain similarities. We adopted the *k*-means clustering algorithm and chose the *k* value, the

number of clusters, based on the silhouette method and elbow rule [27]. The 2 subgroups were generated by applying the 2-means clustering algorithm based on the first five principal components, PC1, PC2, PC3, PC4, and PC5, which were created based on the anthropomorphic attributes. Therefore, the resulting 2 subgroups were very distinct anthropomorphically.

CatBoost Classification and Cross-Validation

We modeled our classifier using the CatBoost model, which is a gradient boosting decision tree-based classifier. The CatBoost model was introduced by Yandex in 2017 and is known to be more accurate for categorical variables than other well-known gradient boosting algorithms such as XGBoost (The XGBoost Contributors), LightGBM (Microsoft Corporation), and GBM (H2O) [28].

To prevent the classification model from overfitting or underfitting, we used stratified *K*-fold cross-validation to train and test the classification models and evaluate the model performance. We first randomly divided the sample into stratified *K*-folds where the folds were formed by preserving the percentage of PCOS-positive in the sample. We used 1-fold as a validation (hold out) set and the remaining (*K*-1) folds for model training.

In the patient model, we used only noninvasive variables (anthropomorphic, symptom, and given variables) for classification. This prediction tool can be used by potential patients or other users who do not have medical testing available. As we do not use the complete set of predictor variables in this patient model, lower accuracy is unavoidable. The provider model uses all predictor variables. This model is expected to aid clinical providers or other users who have access to the patients' medical test results for the prediagnosis or self-diagnosis of PCOS status with higher accuracy. A subgroup study was applied to both models, as summarized in Table 1. We used 5-fold cross-validation with 3 iterations for the entire data set analysis (with a sample size of 526) and 3-fold cross-validation with 5 iterations for the subgroup analysis (with the sample sizes of 287 for Subgroup 1 and 239 for Subgroup 2) to achieve a comparable number of subjects in each fold and averaged the errors over the total 15 folds in both analyses for comparison.

To evaluate the performance of the models, we used four error metrics: accuracy, sensitivity, precision, and F_1 score. Accuracy is the most intuitive and common performance measure, and it is the ratio of correctly predicted observations to the total observations. Sensitivity is the ratio of correctly predicted positive observations to all true positive observations in the class. Precision is the ratio of correctly predicted positive observations to the total number of predicted positive observations. Finally, the F_1 score is the weighted average of the sensitivity and precision. Therefore, it considers both false positives and false negatives. The F_1 score is more useful than the accuracy when there is an uneven class distribution.

XSL·FO

RenderX

Table 1. Proposed models and the difference in method.

	Model description	Predictor variables
Patient model		
Subgroup model	Different model for each subgroup	Only noninvasive variables
One model to all	One model for the entire data set	Only noninvasive variables
Provider model		
Subgroup model	Different model for each subgroup	All variables
One model to all	One model for the entire data set	All variables

Shapley Additive Explanations Values and Feature Importance

The SHAP (Shapley Additive Explanations) values proposed by Lundberg and Lee [29] were used to examine how a single feature affects the output of the model by measuring the change in log odds. The SHAP values can be considered as credit values that are optimally allocated with local explanations using the classic Shapley values in game theory. The SHAP values were first calculated for each variable for each subject in the entire data set and averaged over the sample subjects. SHAP values farther away from 0 have a greater impact on the model output, either positively or negatively. The corresponding total variable importance was calculated by averaging the absolute SHAP values for the sample subjects. This variable importance depicts the magnitude of the total impact of each variable on the model output.

All statistical analyses were performed using R statistical software (version 4.0.3; R Foundation for Statistical Computing). Figure 1 illustrates the subgroup study and summarizes the statistical procedures in a flowchart.

Figure 1. Illustration of the subgroup study and the flowchart of data analysis. PCA: Principal Component Analysis; PCOS: polycystic ovary syndrome.
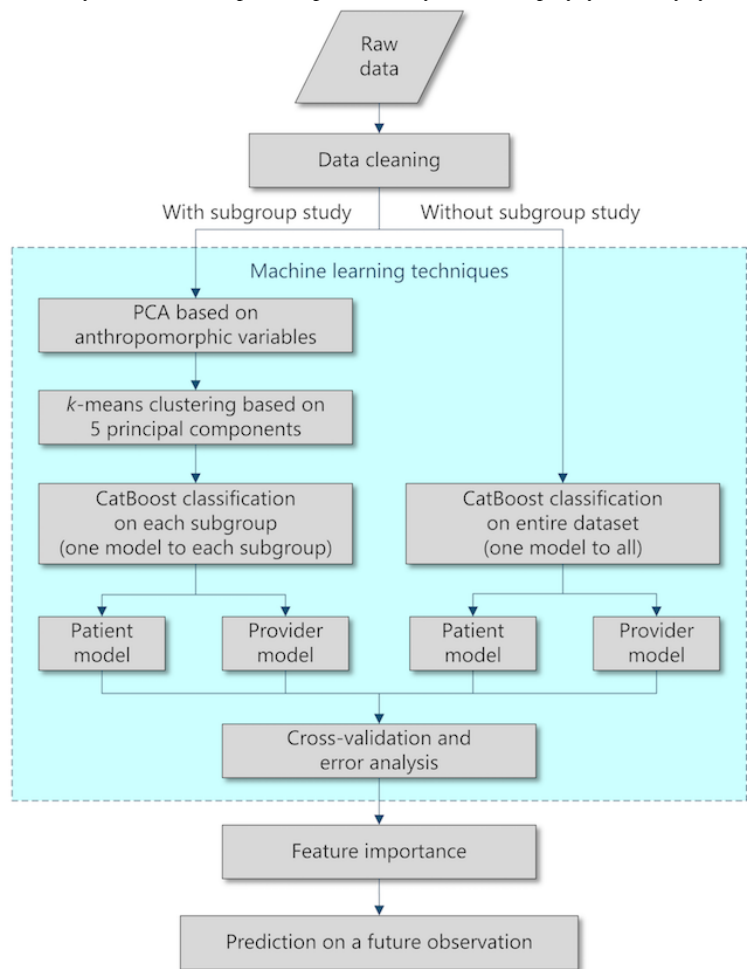
Results

Attributes

We first examined the differences between the PCOS-positive group and the PCOS-negative group to provide an overview of the data set. We used the Anderson-Darling test and found that none of the quantitative variables were normally distributed; therefore, we summarized those variables with median and IQR. For categorical variables, we used percentage. The Wilcoxon test was used to compare the medians for quantitative attributes, and the chi-square test was used to compare the proportions of qualitative attributes between the groups. The corresponding *P* values are presented in the last column of Table 2. Highly significant differences with *P*<.001 in variables were found in

BMI, hip circumference, waist circumference, and body weight among the anthropomorphic variables, all symptom variables, AMH, number of antral follicles in the left ovary and number of antral follicles in the right ovary among the test result variables, and age and fast food consumption among the given variables. These results are consistent with those of the previous studies. Anthropomorphic variables have often been linked to PCOS [30]. Symptoms and AMH have also been reported as important markers for the diagnosis of PCOS [31]. As the number of antral follicles in the left ovary and number of antral follicles in the right ovary are used in the diagnosis of PCOS, the significant differences in these variables between the 2 groups were significant. Finally, age and diet habits were found to be important factors for PCOS [32,33]. Figures 2-5 display the group comparison of these variables in detail.

Table 2. Anthropomorphic, symptom, test result, and given variables. Comparison between the PCOS[a]-positive group and the PCOS-negative group with *P* values (N=526).

	All women (N=526)	PCOS-positive size (n=170)	PCOS-negative size (n=356)	*P* value[b]
Anthropomorphic variables, median (IQR)				
BMI	24.27 (21.69 to 26.66)	25.1 (26.25 to 33)	23.61 (21.37 to 26.13)	<.001
Height (cm)	156 (152 to 160)	158 (152 to 161)	156 (152 to 160)	.12
Hip (inches)	38 (36 to 40)	39 (36 to 42)	38 (36 to 40)	<.001
Waist circumference (inches)	34 (52 to 65)	35 (32 to 37)	34 (31.75 to 35)	<.001
Waist-to-hip ratio	0.89 (0.88 to 0.93)	0.89 (0.86 to 0.93)	0.89 (0.86 to 0.93)	.91
Weight (kg)	59 (52 to 65)	62 (55 to 70)	58 (52 to 64)	<.001
Symptom variables				
Acanthosis nigricans (%)	158 (30)	105 (61.8)	53 (14.9)	<.001
Acne (%)	256 (48.7)	118 (69.4)	138 (38.8)	<.001
Hair loss (%)	236 (44.9)	97 (57.1)	139 (39)	<.001
Hirsutism (%)	143 (27.2)	22 (12.9)	46 (12.9)	<.001
Irregular menstrual cycle (%)	146 (27.8)	91 (53.5)	55 (15.5)	<.001
Length of menstrual cycle (days), median (IQR)	5 (5 to 5)	5 (3 to 5)	5 (5 to 6)	<.001
Weight gain (%)	199 (37.8)	117 (68.8)	82 (23)	<.001
Test result variables, median (IQR)				
AMH[c,d] (ng/mL)	1.31 (0.70 to 1.92)	1.74 (0.88 to 2.32)	1.16 (0.65 to 1.66)	<.001
Antral follicles in left ovary	5 (3 to 9)	10 (7 to 12)	4 (2 to 6)	<.001
Antral follicles in right ovary	6 (3 to 10)	11 (8 to 13)	4 (2 to 7)	<.001
Average follicle size left ovary (mm)	15 (13 to 18)	16 (14 to 18)	15 (13 to 18)	.02
Average follicle size right ovary (mm)	16 (13 to 18)	16 (14 to 18)	15.5 (13 to 18)	.046
Diastolic BP (mm Hg)	80 (70 to 80)	80 (70 to 80)	80 (70 to 80)	.69
Endometrium thickness (mm)	8.5 (7 to 9.8)	8.9 (7.6 to 10)	8.3 (7 to 9.5)	.006
FSH[c,e] (mIU/mL)	1.58 (1.2 to 1.85)	1.51 (1.17 to 1.75)	1.61 (1.25 to 1.88)	.007
FSH to LH[f] ratio[c]	0.77 (0.35 to 1.35)	0.69 (0.07 to 1.14)	0.86 (0.44 to 1.41)	.002
Glycated hemoglobin level (g/100 ml)	11 (10.5 to 11.7)	11.05 (10.7 to 11.9)	11 (10.5 to 11.5)	.03
HCG[g] I[c] (mIU/mL)	2.96 (0.688 to 5.7)	4.25 (0.69 to 5.85)	2.68 (0.69 to 5.62)	.10
HCG II[c] (mIU/mL)	0.69 (0.688 to 4.61)	0.69 (0.69 to 4.63)	0.69 (0.69 to 4.59)	.94
LH (mIU/mL)	2.25 (1.03 to 3.68)	2.23 (1.03 to 4.31)	2.3 (1.03 to 3.6)	.25
PRG[c,h] (ng/mL)	−1.14 (−1.39 to −0.78)	−1.14 (−1.39 to −0.84)	−1.17 (−1.39 to −0.78)	.54
PRL[c,i] (ng/mL)	3.09 (2.68 to 3.39)	3.13 (2.64 to 3.41)	3.06 (2.69 to 3.39)	.73
Pulse rate (beats/min)	72 (72 to 74)	72.5 (72 to 74)	72 (72 to 74)	.002
Random glucose test (mg/100 mL)	100 (92 to 107)	100 (92 to 107)	97.5 (92 to 108)	.32
Respiratory rate (breaths/min)	18 (18 to 20)	20 (18 to 20)	18 (18 to 20)	.28
Systolic BP[j] (mm Hg)	110 (110 to 120)	110 (110 to 120)	110 (110 to 120)	.73
Thyrotropin[c] (mIU/L)	0.82 (0.39 to 1.27)	0.84 (0.39 to 1.26)	0.78 (0.39 to 1.27)	.69
Vitamin D3 (ng/mL)	25.95 (20.73 to 34.48)	25.38 (19.3 to 33.58)	26.3 (21.28 to 35.8)	.19
Given variables				
Age (years), median (IQR)	31 (27 to 35)	29 (26.25 to 33)	32 (28 to 36)	<.001

	All women (N=526)	PCOS-positive size (n=170)	PCOS-negative size (n=356)	P value[b]
Blood type (%)				.97
A+	105 (20)	33 (19.4)	72 (20.2)	
A−	13 (2.5)	4 (2.4)	9 (2.5)	
B+	130 (24.7)	40 (23.5)	90 (25.3)	
B−	16 (3)	6 (3.5)	10 (2.8)	
O+	200 (38)	63 (37.1)	137 (38.5)	
O−	19 (3.6)	8 (4.7)	11 (3.1)	
AB+	41 (7.8)	15 (8.8)	26 (7.3)	
AB-	2 (0.4)	1 (0.6)	1 (0.3)	
Married (years), median (IQR)	7 (4 to 10)	6 (3 to 9)	7 (4 to 11)	.001
Number of abortions (%)				.61
0	425 (80.8)	141 (82.9)	284 (79.8)	
1	68 (12.9)	21 (12.4)	47 (13.2)	
2 and above	33 (6.3)	8 (4.7)	25 (7)	
Fast food consumption (%)	270 (51.3)	134 (78.8)	136 (38.2)	<.001
Pregnant (%)	204 (38.8)	63 (37.1)	141 (39.6)	.64
Regular exercise (%)	127 (24.1)	48 (28.2)	79 (22.2)	.16

[a]PCOS: polycystic ovary syndrome.

[b]P values were calculated by the Wilcoxon test for median comparison and chi-square test for proportion comparison.

[c]Log-transformed variables.

[d]AMH: anti-Müllerian hormone.

[e]FSH: follicle-stimulating hormone.

[f]LH: luteinizing hormone.

[g]HCG: human chorionic gonadotropin.

[h]PRG: progrestrone.

[i]PRL: prolactin.

[j]BP: blood pressure.

Figure 2. Anthropomorphic variable comparison between the polycystic ovary syndrome (PCOS) positive group and the PCOS-negative group with *P*<.001.
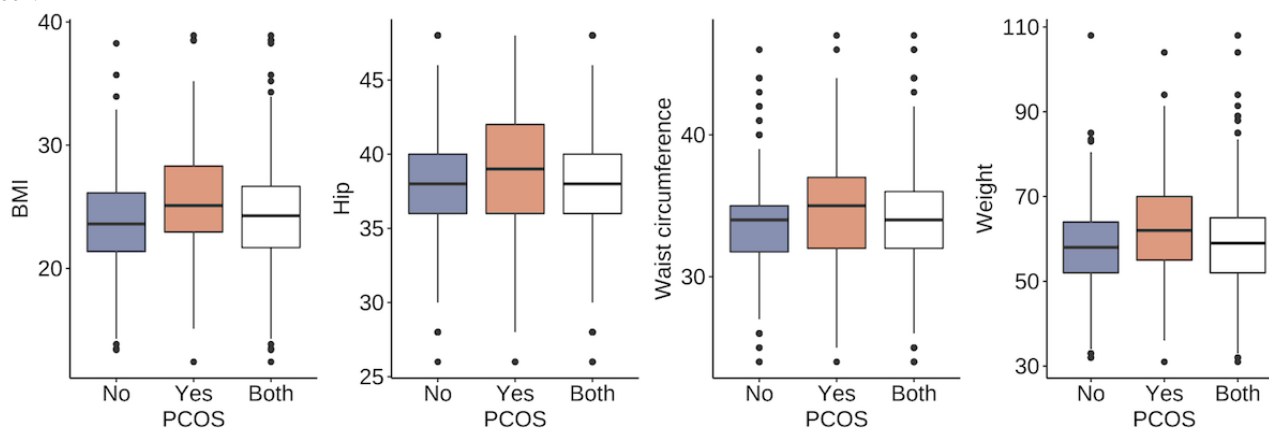
Figure 3. Symptom variable comparison between the polycystic ovary syndrome (PCOS) positive group and the PCOS-negative group with $P<.001$.
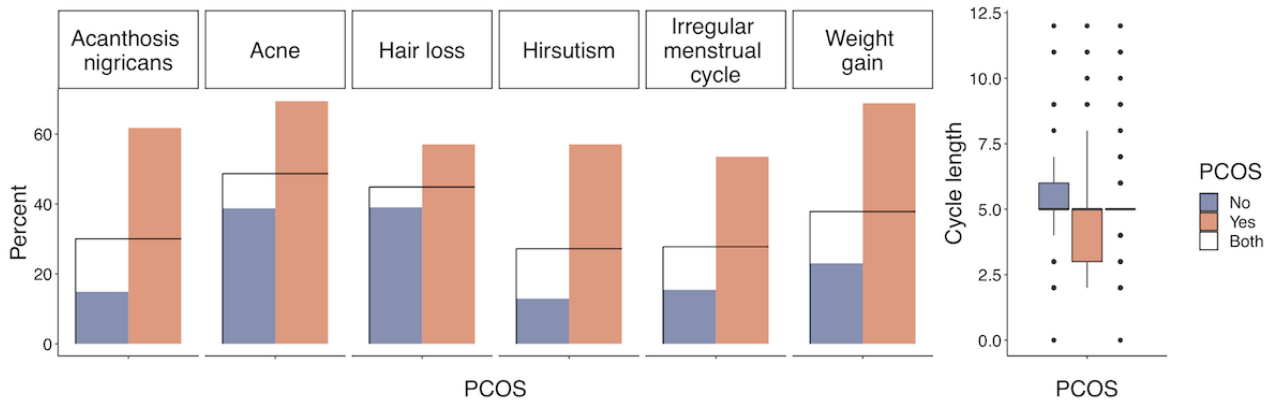
Figure 4. Test result variable comparison between the polycystic ovary syndrome (PCOS) positive group and the PCOS-negative group with $P<.001$. AMH: anti-Müllerian hormone.
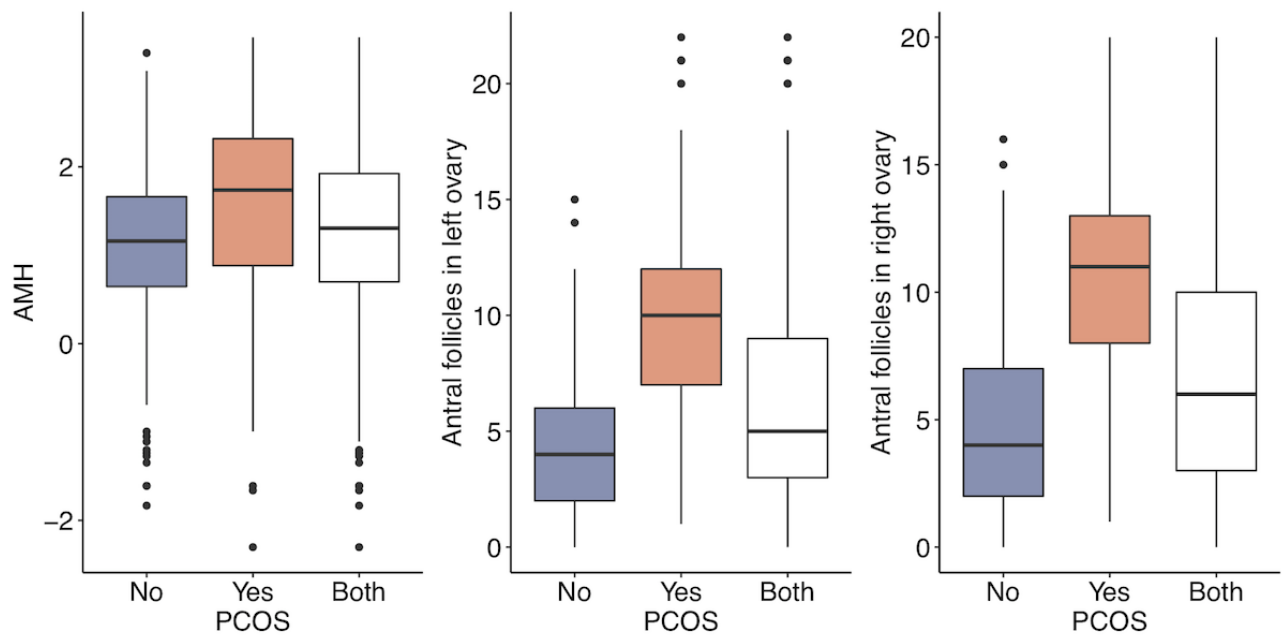
Figure 5. Given variable comparison between the polycystic ovary syndrome (PCOS) positive group and the PCOS-negative group with *P*<.001.
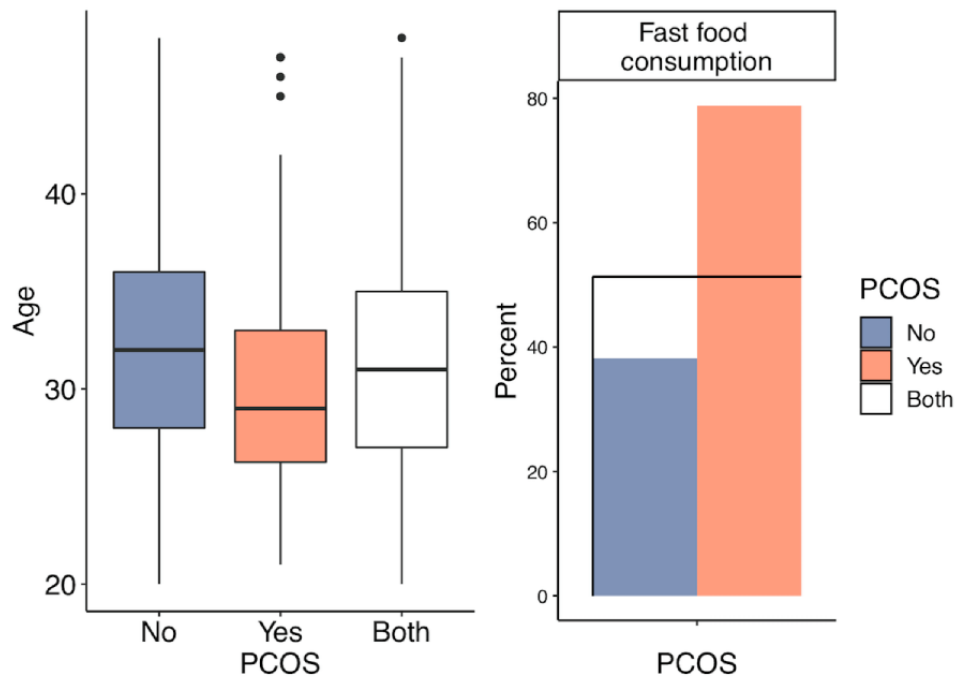
Feature Extraction and Clustering

For the subgroup study, we used the first 5 principal components constructed based on the anthropomorphic attributes and then applied *k*-means clustering using the 5 principal components to classify the 526 women into 2 subgroups. Figure 6 shows the correlation among the anthropomorphic variables. BMI, hip circumference, waist circumference, and body weight are highly and positively correlated, and height and waist-to-hip ratio are negatively correlated to the rest of variables. This can be also observed in PCA. The biplot in Figure 7 and Table 3 explain how each anthropomorphic variable contributes to the individual principal components. For example, as BMI, hip circumference, waist circumference, and body weight are closely correlated as depicted in Figure 6, these variables largely contribute to the first principal component (the horizontal axis, PC1) with high loadings in Figure 7 and Table 3. This newly constructed PC1 accounts for or *explains* 53.5% of the overall variability in Figure 7, and it is the first principal axis in the direction where the data points vary the most. The waist-to-hip ratio variable highly contributes to the second principal component (the vertical axis, PC2) with the second largest proportion (15.8%) of variance explained in Figure 7 and Table 3. The number of

principal components is chosen based on the amount of variance explained by using the principal components. In Figure 8, we observe that the proportion of variance explained drops at 5 principal components and we choose the first 5 principal components for clustering in the next step.

For *k*-means clustering, the optimal number *(k)* is determined based on the silhouette method and the elbow rule in Figure 9, which shows the average silhouette width versus the number of clusters. As indicated in Figure 9, we chose *k*=2, where the curve shows a sharp kink (the elbow rule). The motivation of the 2-subgroup study also originated from the idea of predicting lean PCOS and obese PCOS in a more detailed manner. In Figure 10, the 2 subgroups are shown in the PC1-PC2 plane. Subgroup 1 included 287 subjects with 76 PCOS-positive cases and subgroup 2 had 239 subjects with 94 PCOS-positive cases. The mean and SD of BMI for Subgroup 1 were 21.862 kg/m^2 and 2.921 kg/m^2, and those for Group 2 were 27.246 kg/m^2 and 3.183 kg/m^2, respectively. This clustering provides a similar yet more structured subgrouping to the lean and obese PCOS groups, where the lean PCOS group had a BMI <25 kg/m^2 and the obese PCOS group had a BMI of ≥25 kg/m^2 [34].

Figure 6. Correlation between the anthropomorphic variables.
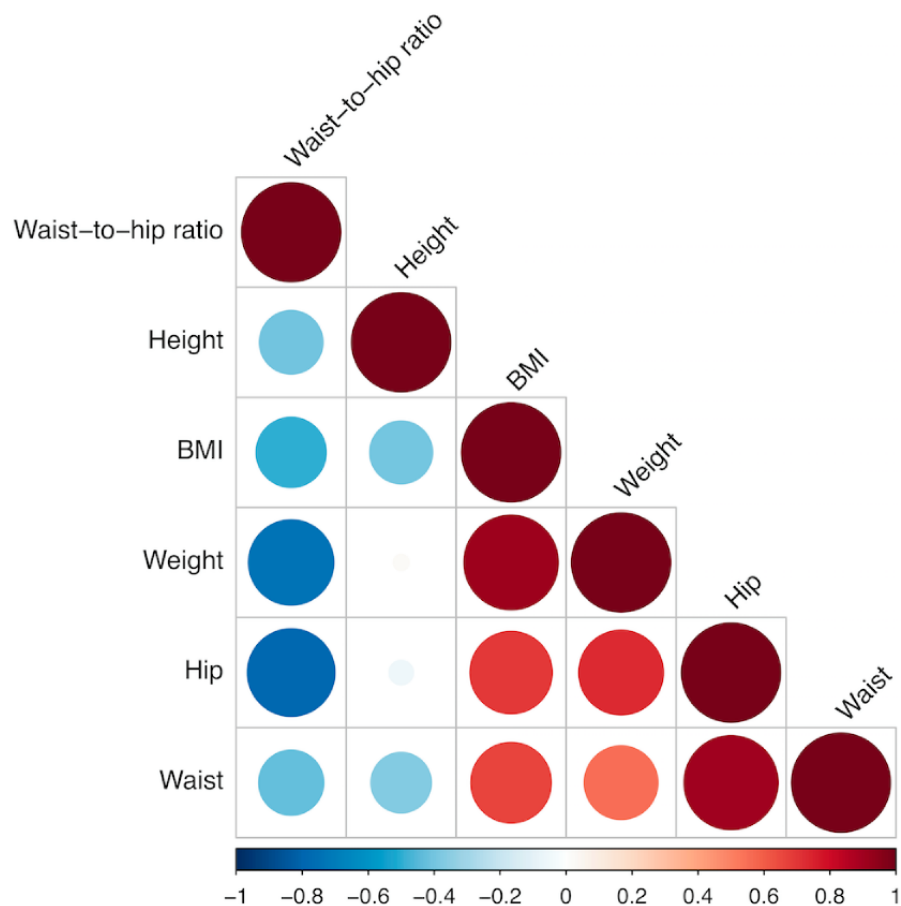
Figure 7. Principal Component Analysis. Biplot of PC1 and PC2 based on the anthropomorphic attributes.
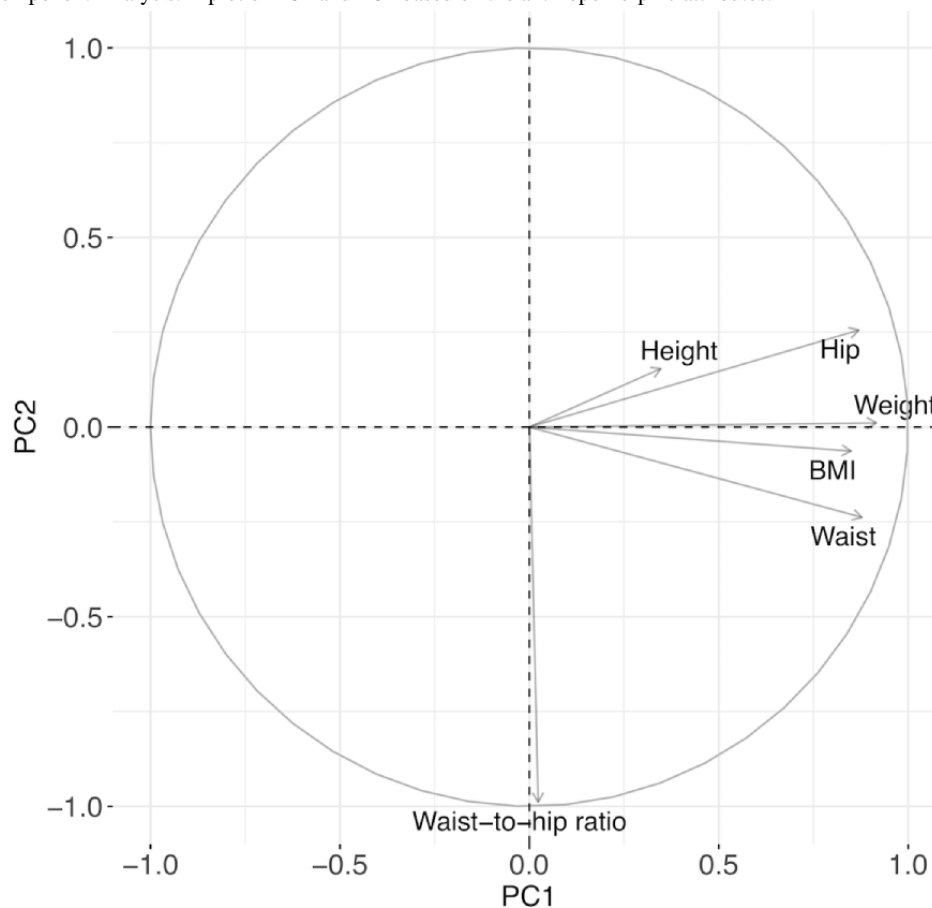
Table 3. Variable loadings[a] related to each principal component.

Anthropomorphic variable	Principal components				
	PC1	PC2	PC3	PC4	PC5
BMI	0.474	−0.060	0.322	−0.509	0.640
Height	0.194	0.145	−0.921	−0.040	0.302
Hip	0.486	0.241	0.119	0.500	0.010
Waist-to-hip ratio	0.013	−0.931	−0.146	0.028	0.003
Waist	0.490	−0.224	0.046	0.510	−0.010
Weight	0.511	0.010	−0.102	−0.477	−0.707

[a]Loading denotes the contribution of the variable to each principal component. Higher absolute value indicates more contribution to the corresponding principal component.

Figure 8. The optimal number of principal components to be used based on the proportion of variance explained.
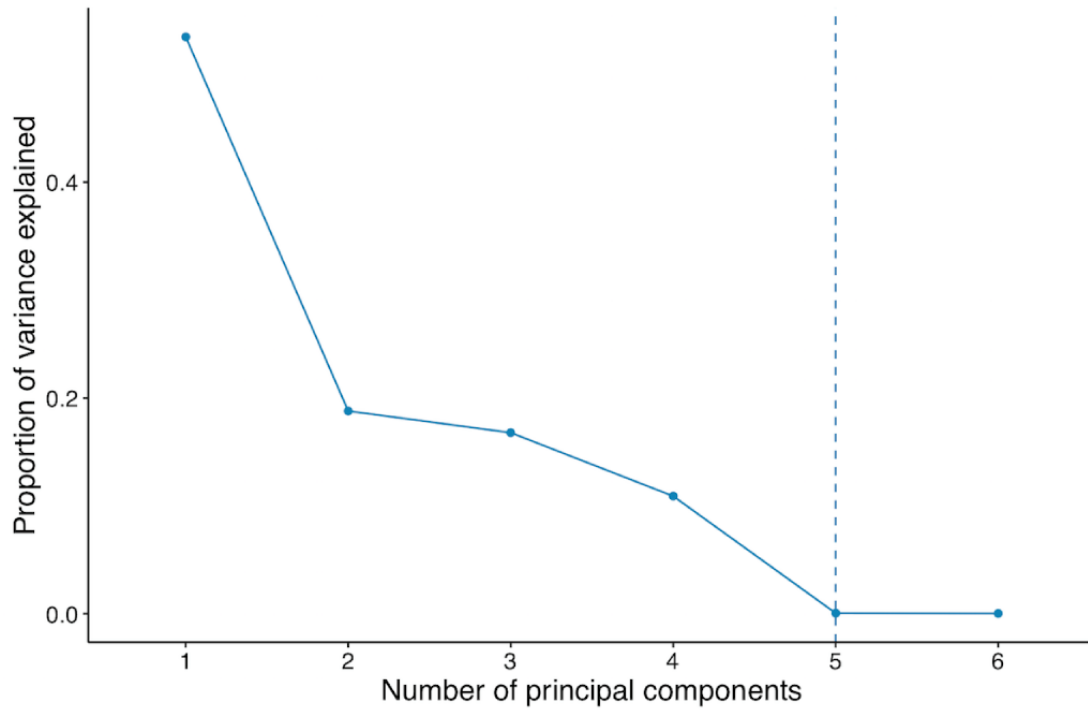
Figure 9. Silhouette method to determine the optimal number of clusters.
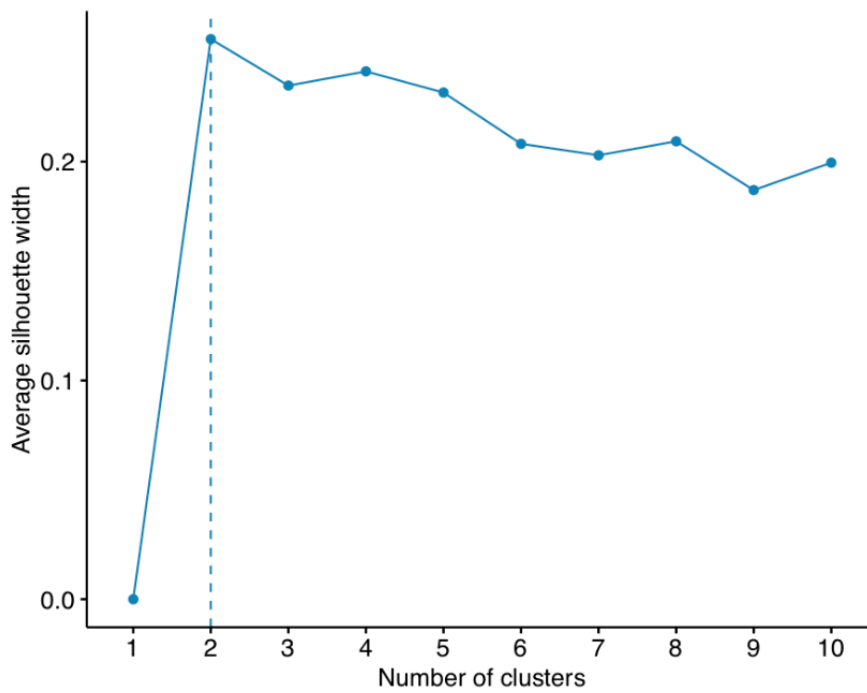
Figure 10. Subgroups on the PC1-PC2 plane after applying *k*-means clustering using the first 5 principal components based on the anthropomorphic attributes.
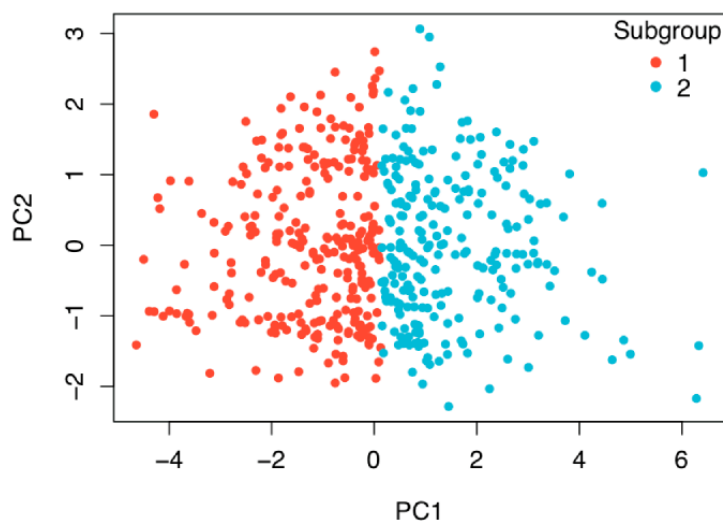
In Table 4, these 2 subgroups are compared in great detail. The same statistical tests used in Table 2 were used for this comparison. These subgroups were generated by using the clustering algorithm based on the first 5 principal components of the anthropomorphic variables as discussed above. Consequently, there were significant differences in all anthropomorphic variables except waist-to-hip ratio. Otherwise, there were highly significant differences with *P*<.001 in hirsutism, irregular menstrual cycle, and weight gain among the symptom variables. We visually compare these variables in Figures 11 and 12.

Table 4. Characteristics of 2 subgroups classified by *k*-means clustering based on the 5 principal components of the anthropomorphic variables.

	Subgroup 1; n=287, PCOS[a]-positive, n=76	Subgroup 2; n=239, PCOS-positive, n=94	*P* value[b]
Anthropomorphic variables, median (IQR)			
BMI (kg/m^2)	22.15 (20.29 to 23.9)	26.75 (25.1 to 28.98)	<.001
Height (cm)	154 (152 to 158)	158 (154 to 1620)	<.001
Hip (in)	36 (34 to 38)	40 (39 to 42)	<.001
Waist circumference (in)	32 (30 to 34)	36 (35 to 38)	<.001
Waist-to-hip ratio	0.89 (0.85 to 0.93)	0.9 (0.86 to 0.93)	.24
Weight (kg)	53 (50 to 56)	66 (62 to 72.15)	<.001
Symptom variables			
Acanthosis nigricans (%)	76 (26.5)	82 (34.3)	.06
Acne (%)	141 (49.1)	115 (48.1)	.89
Hair loss (%)	123 (42.9)	113 (47.3)	.35
Hirsutism (%)	58 (20.2)	85 (35.6)	<.001
Irregular menstrual cycle (%)	60 (20.9)	137 (57.2)	<.001
Length of menstrual cycle (days), median (IQR)	5 (5 to 5)	5 (4 to 6)	.73
Weight gain (%)	59 (20.6)	140 (58.6)	<.001
Test result variables, median (IQR)			
AMH[c,d] (ng/mL)	1.34 (0.71 to 1.92)	1.29 (0.69 to 1.92)	.92
Antral follicles in left ovary	5 (3 to 8)	6 (3 to 9)	.06
Antral follicles in right ovary	5 (3 to 9)	7 (3 to 10)	.06
Average follicle size left ovary (mm)	15 (13.5 to 18)	15 (13 to 18)	.27
Average follicle size right ovary (mm)	16 (13 to 18)	16 (13 to 18)	.28
Diastolic BP[f] (mm Hg)	80 (70 to 80)	80 (80 to 80)	.02
Endometrium thickness (mm)	8.5 (7 to 10)	8.5 (7 to 9.6)	.39
FSH[c,f] (mIU/mL)	1.6 (1.18 to 1.87)	1.53 (1.21 to 1.83)	.34
FSH to LH[g] ratio[c]	0.73 (0.38 to 1.38)	0.67 (0.29 to 1.35)	.93
Glycated hemoglobin (g/100 ml)	11 (10.6 to 11.75)	11 (10.5 to 11.7)	.69
HCG I[c,h] (mIU/mL)	3.77 (0.69 to 5.91)	2.3 (0.69 to 5.44)	.02
HCG II[c] (minus/mL)	0.69 (0.69 to 4.51)	0.69 (0.69 to 4.61)	.89
LH (maul/mL)	2.34 (1.04 to 3.84)	2.15 (1.03 to 3.46)	.51
PRG[c,i] (ng/mL)	−1.2 (−1.39 to −0.8)	−1.11 (−1.39 to −0.78)	.18
PRL[c,j] (ng/mL)	3.14 (2.75 to 3.14)	3 (2.62 to 3.36)	.02
Pulse rate (beats/min)	72 (72 to 74)	72 (72 to 74)	.61
Random glucose test (mg/100 mL)	98 (91.5 to 106)	100 (92 to 108)	.17
Respiratory rate (breaths/min)	18 (18 to 20)	18 (18 to 20)	.79
Systolic BP (mm Hg)	110 (110 to 120)	120 (110 to 120)	.90
Thyrotropin[c] (mIU/L)	0.79 (0.37 to 1.29)	0.83 (0.41 to 1.22)	.98
Vitamin D3 (ng/mL)	26 (21.26 to 34.3)	25.69 (20.05 to 34.9)	.56
Given variables			
Age (years), median (IQR)	31 (28 to 35)	31 (27 to 35)	.33

	Subgroup 1; n=287, PCOS[a]-positive, n=76	Subgroup 2; n=239, PCOS-positive, n=94	P value[b]
Blood type (%)			.007
A+	59 (20.6)	46 (19.3)	
A−	6 (2.1)	7 (2.9)	
B+	79 (27.5)	51 (21.3)	
B−	12 (4.2)	4 (1.7)	
O+	99 (34.5)	101 (42.3)	
O−	5 (1.7)	14 (5.9)	
AB+	27 (9.4)	14 (5.9)	
AB−	0 (0)	2 (0.8)	
Married (years), median (IQR)	7 (4 to 10)	7 (4 to 10)	.97
Number of abortions (%)			.31
0	239 (83.3)	186 (77.8)	
1	35 (12.2)	33 (13.8)	
2 and above	13 (4.5)	20 (8.4)	
Fast food consumption (%)	132 (46)	138 (57.7)	.009
Pregnant (%)	117 (40.8)	87 (36.4)	.35
Regular exercise (%)	83 (28.9)	44 (18.4)	.007

[a]PCOS: polycystic ovary syndrome.

[b]P values were calculated by the Wilcoxon test for median comparison and chi-square test for proportion comparison.

[c]Log-transformed variables.

[d]AMH: anti-Müllerian hormone.

[e]BP: blood pressure.

[f]FSH: follicle-stimulating hormone.

[g]LH: luteinizing hormone.

[h]HCG: human chorionic gonadotropin.

[i]PRG: progrestrone.

[j]PRL: prolactin.

Figure 11. Anthropomorphic variable comparison between subgroup 1 and subgroup 2 with $P<.001$.
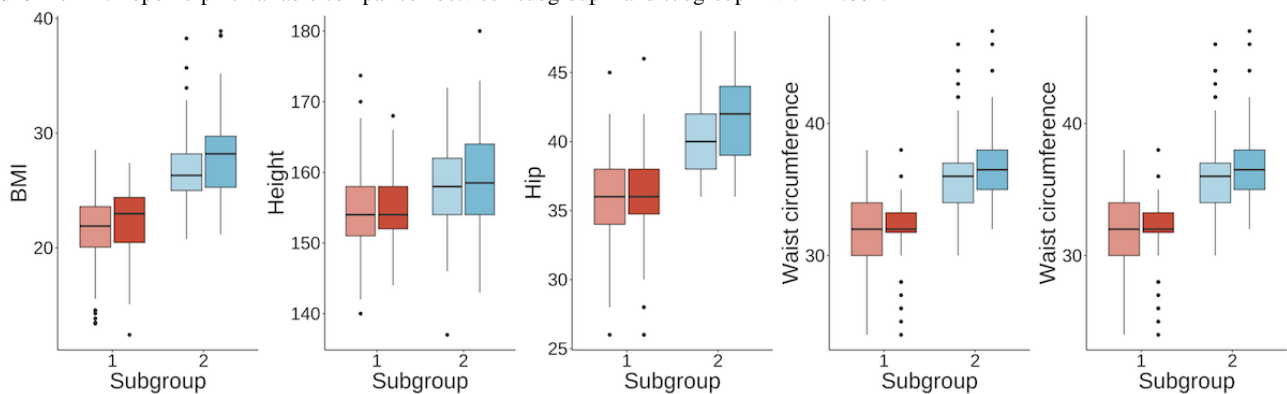
Figure 12. Symptom variable comparison between subgroup 1 and subgroup 2 with *P*<.001. PCOS: polycystic ovary syndrome.
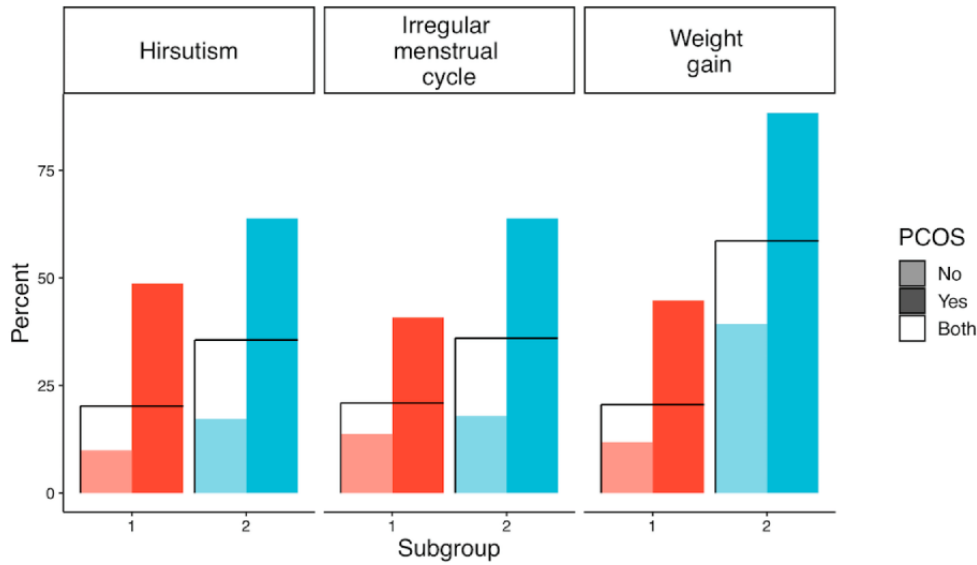
Performance Evaluation

The prediction results from the 4 different models in Table 1 are compared in Tables 5 and 6. We used four different error metrics: accuracy, sensitivity, precision, and F_1 score. Table 5 shows the results of the patient models with and without using subgroups. As expected, the subgroup models had overall lower performance in predicting PCOS status. With subgroups, the prediction in the lower BMI group (subgroup 1) was slightly higher than that of the other subgroup. This might be because

women with higher BMI are more likely to have other complications [14,15], causing more difficulty in predicting the status of PCOS. In Table 6, we present a comparison of the error metrics for the provider models with and without subgroups. In the provider models, we predict the status of PCOS using the complete set of predictor variables, including the noninvasive variables as well as the test result variables. Therefore, the model prediction is overall better than the patient models, where we use only noninvasive predictors.

Table 5. Summary of averaged accuracy, sensitivity, precision, and F_1 score for one model to each subgroup (with subgroups) versus one model to all (without subgroups) in the patient model, using only noninvasive predictor variables in the models.

Metrics	Patient model with subgroup		Patient model without subgroup
	Subgroup 1	Subgroup 2	
Accuracy	0.815	0.810	0.825
Sensitivity	0.837	0.827	0.851
Precision	0.931	0.870	0.900
F_1 score	0.880	0.846	0.874

Table 6. Summary of averaged accuracy, sensitivity, precision, and F_1 score for one model to each subgroup (with subgroups) versus one model to all (without subgroups) in the provider model, using all predictor variables in the models.

Metrics	Provider model with subgroup		Provider model without subgroup
	Subgroup 1	Subgroup 2	
Accuracy	0.898	0.875	0.901
Sensitivity	0.907	0.862	0.909
Precision	0.960	0.945	0.950
F_1 score	0.932	0.901	0.928

Feature Analysis

The SHAP values and the corresponding feature importance are examined in this subsection. First, we present patient models based only on noninvasive predictor variables. Figures 13-15 show the total variable importance in the first column for all

noninvasive predictor variables, and the SHAP values for the 12 most important variables are graphed in the second column for the patient models. Figure 13 shows the result for subgroup 1 and Figure 14 shows the result for subgroup 2. In both subgroups, all symptom variables except hair loss and the two given variables, fast food consumption and age, are important

features. The major difference between the subgroups is that acanthosis nigricans (*Darkening* in the plot) is the most important feature in subgroup 1, and weight gain is the most important feature in subgroup 2. Another interesting finding is that exercise is more important in subgroup 1, and irregular menstrual cycle is more prominent in subgroup 2 than in the other subgroup. In Figure 15, the results from one model to all (without subgrouping) in the patient model are presented. All symptom variables except hair loss and the two given variables, fast food consumption and age, have high importance values as well.

Figure 13. Feature importance for all variables and SHAP (Shapley Additive Explanations) values for the 12 most important features of subgroup 1 in the patient subgroup model including only noninvasive predictor variables.
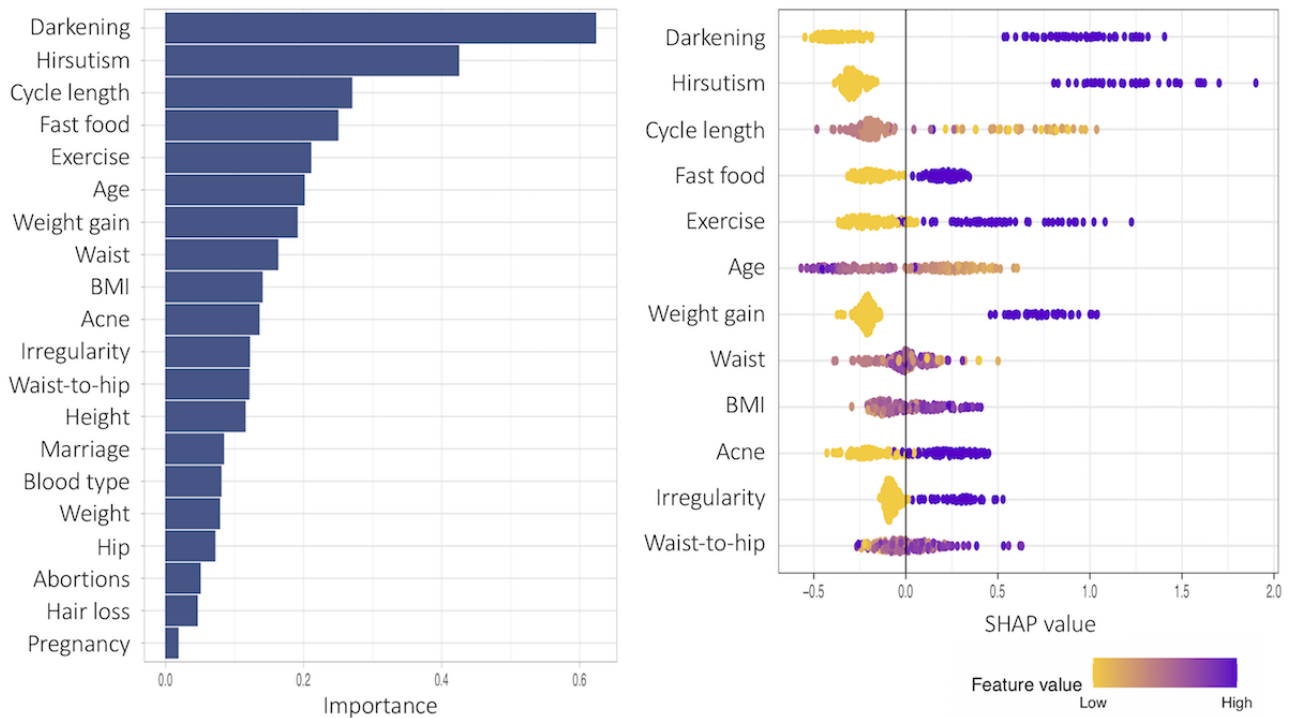
Figure 14. Feature importance for all variables and SHAP (Shapley Additive Explanations) values for the 12 most important features of subgroup 2 in the patient subgroup model including only noninvasive predictor variables.
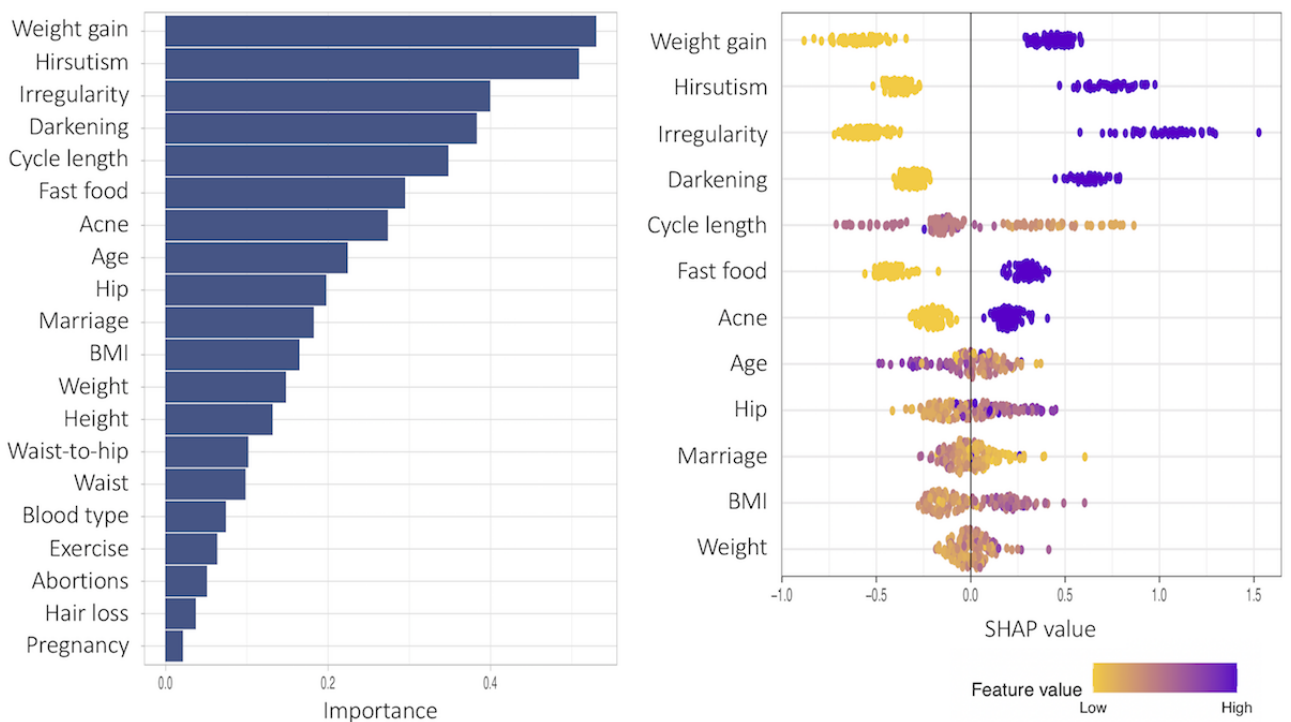
Figure 15. Feature importance for all variables (with the middle importance variables omitted) and the SHAP (Shapley Additive Explanations) values for the 12 most important features of the one to all model (without subgroups) in the patient model.
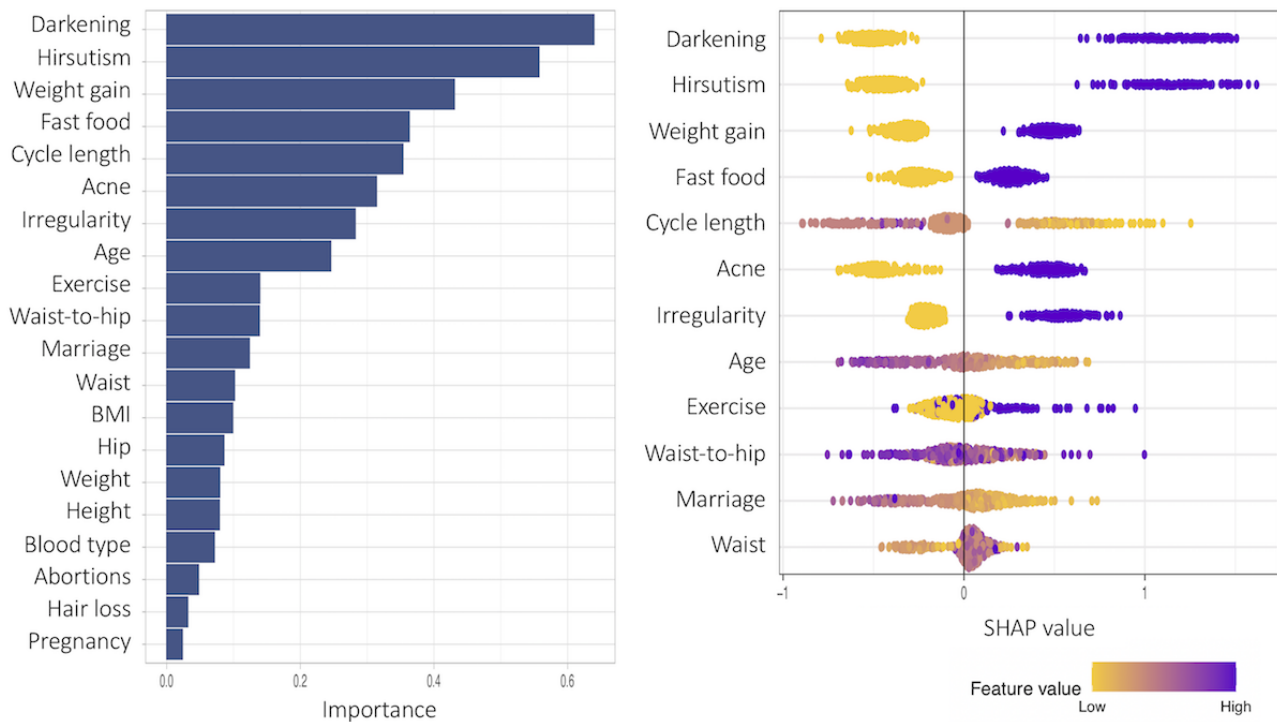
The SHAP values provide detailed local behavior in terms of feature importance. Most binary variables have distinct positive and negative SHAP values. For example, the SHAP values for the darkening variable, top-ranked in subgroup 1 in Figure 13, are clustered on the negative side (yellow dots in the plot) and on the positive side (purple dots in the plot), and these 2 clusters are far away from zero. The yellow dots represent lower values in the darkening variable (Darkening=No) and these variable values negatively affect the model output, that is, negative change in log odds in the model output, whereas the purple dots show higher values in the darkening variable (Darkening=Yes), and these variable values positively affect the model output. However, the higher feature values (Darkening=Yes) in purple have a stronger impact on the model output than the lower feature values (Darkening=No) because the purple cluster is farther away from zero than the yellow cluster.

In Figures 16-18, we repeat the same process in the provider models. In all 3 figures, the number of antral follicles in the right ovary was most highly ranked in terms of feature importance. Otherwise, the number of antral follicles in the left ovary, all symptom attributes except hair loss, fast food consumption, and AMH are relatively more important than other variables, which are commonly observed in all 3 models. In comparison between subgroup 1 and subgroup 2, pulse rate and respiratory rate are more important features in subgroup 1 in Figure 16, and AMH is a more important feature (also note that the SHAP values of AMH are more spread out) in subgroup 2 than in the other subgroup, as shown in Figure 17. Regarding pulse rate and respiratory rate being important in subgroup 1, heart rate variability in normal weight women with PCOS has been subject to debate in the literature [35,36].

Figure 16. Feature importance for all variables (with the middle importance variables omitted) and SHAP (Shapley Additive Explanations) values for the 12 most important features of subgroup 1 in the provider subgroup model including all predictor variables. AMH: anti-Müllerian hormone; PRG: progrestrone; PRL: prolactin.
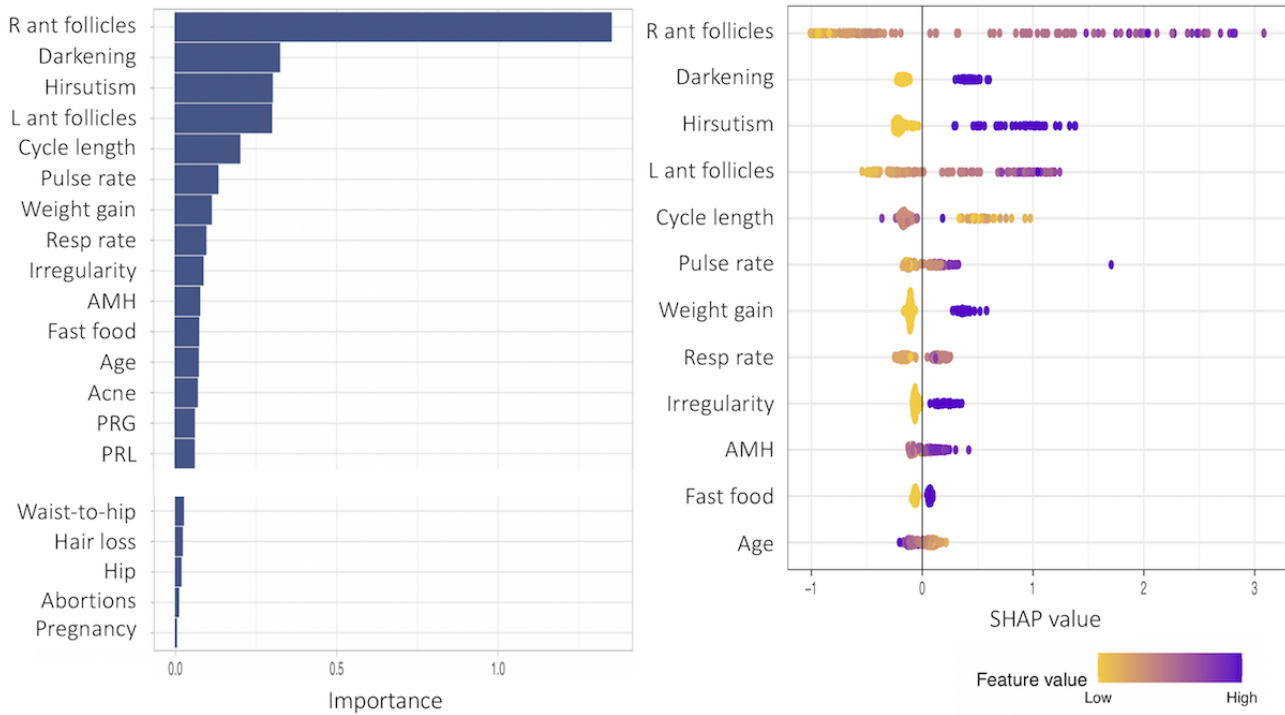
Figure 17. Feature importance for all variables (with the middle importance variables omitted) and SHAP (Shapley Additive Explanations) values for the 12 most important features of subgroup 2 in the provider subgroup model including all predictor variables. AMH: anti-Müllerian hormone; BP: blood pressure; FSH: follicle-stimulating hormone; LH: luteinizing hormone; PRL: prolactin.
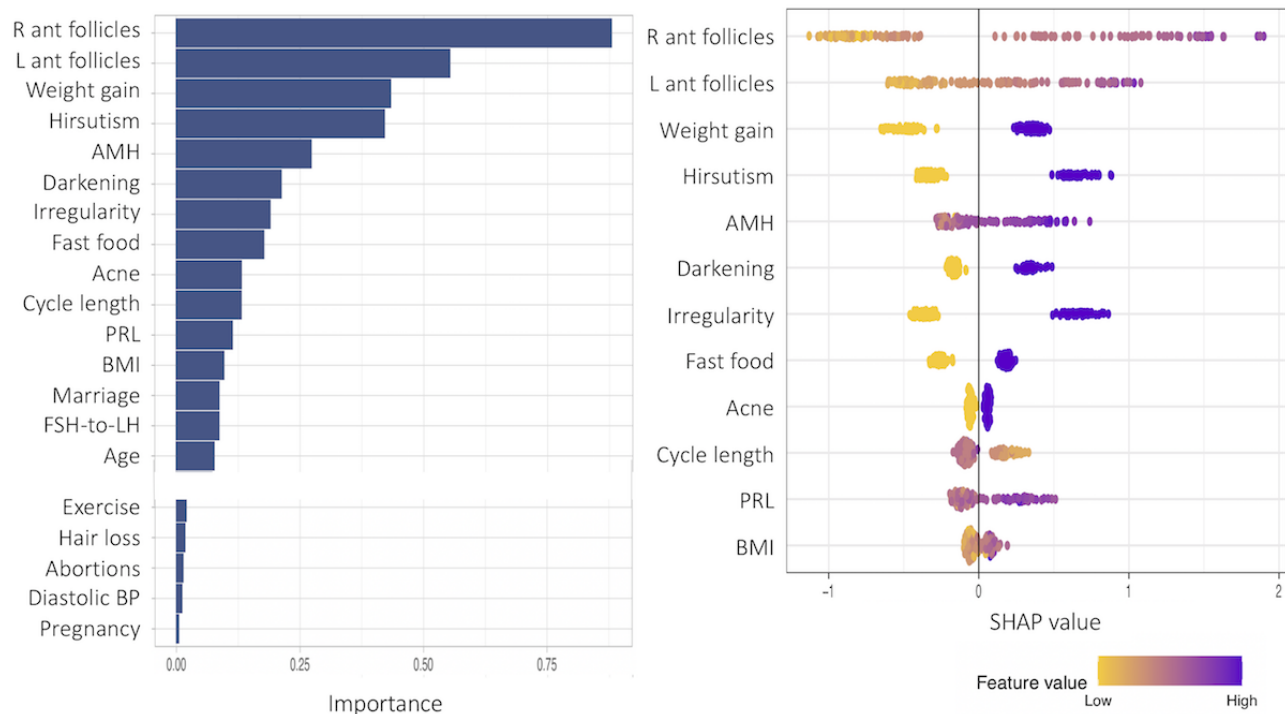
Figure 18. Feature importance for all variables (with the middle importance variables omitted) and the SHAP (Shapley Additive Explanations) values for the 12 most important features of the one to all model (without subgroups) in the provider model. AMH: anti-Müllerian hormone; BP: blood pressure; FSH: follicle-stimulating hormone; LH: luteinizing hormone; PRL: prolactin.
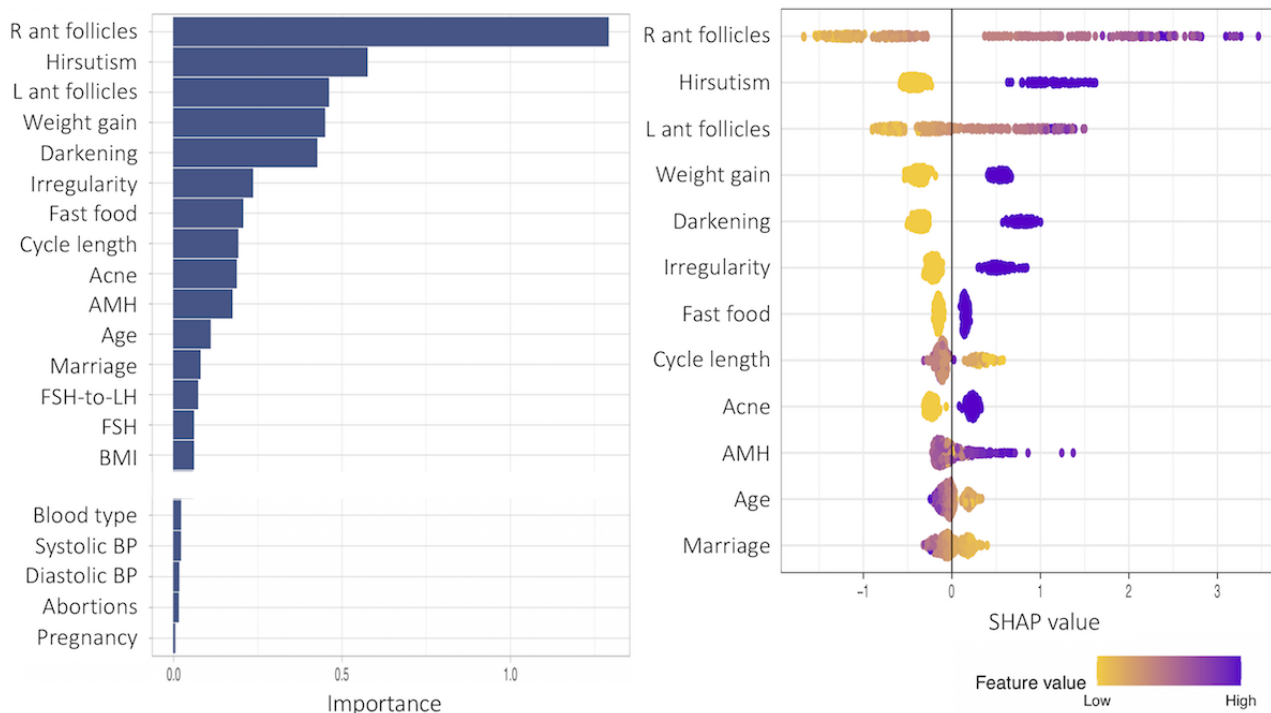
Discussion

Principal Findings

PCOS is a common, chronic, yet underrecognized female hormone disorder. Owing to the complexity of the syndrome, identification and differential diagnosis remain challenging even with widely accepted criteria and tests. Another disparity comes from the gaps in early diagnosis, information, and accessible support that can help women to prevent or manage this life span syndrome more adequately.

The ultimate goal of this study was to develop a conveniently accessible digital platform for pre- or self-diagnosis of PCOS using machine learning techniques such as PCA, k-means clustering algorithm, and CatBoost classifier based on the health measures of 526 female subjects. PCA was adopted to extract features from highly correlated anthropomorphic variables. The 2-means clustering algorithm was used to classify the 526 women into 2 different subgroups based on the first 5 principal components, leading to 2 subgroups with distinct BMIs. The gradient boosting decision tree-based classifier, CatBoost model, was trained and tested, and the prediction (test) error rates were compared between models based on four different error metrics: accuracy, sensitivity, precision, and F_1 score.

We developed 2 types of prediction models targeting different groups of users. One model is for potential patients or other users who do not have medical test results available (the patient model) and the other is for clinical providers or other users who have access to the patients' medical records and test results (the provider model). In each model, we applied a subgroup study to obtain the detailed characteristics of the 2 subgroups in the analysis. For the patient models, the prediction accuracy ranged

from 81% to 81.5% with subgroups and 82.5% without subgroups. For the provider models, the prediction accuracy ranged from 87.5% to 89.8% with subgroups and 90.1% without subgroups.

Feature importance was performed in each model based on the SHAP values and the corresponding total feature importance. In the patient models, all symptom variables other than hair loss, that is, acanthosis nigricans, acne, hirsutism, irregular menstrual cycle, length of menstrual cycle, and weight gain, were important along with fast food consumption and age. For subgroup 1 (with lower BMI), acanthosis nigricans was the strongest marker in the prediction of PCOS status. Exercise was also an important factor in subgroup 1. For subgroup 2 (with higher BMI), weight gain was the top-ranked important marker and irregular menstrual cycle was also more prominent than subgroup 1. For the provider models, the number of antral follicles in the right ovary, the number of antral follicles in the left ovary, and AMH were important in all models in addition to the listed important variables in the patient models: acanthosis nigricans, acne, hirsutism, irregular menstrual cycle, length of menstrual cycle, weight gain, fast food consumption, and age. Another interesting finding was that the pulse rate and respiratory rates were also highly ranked in subgroup 1, which consisted of normal or underweight women with a mean weight of 52.353 kg and a mean BMI of 21.862 kg/m^2.

The proposed prediction models based on available health measures are expected to provide women with a simpler and quicker access to a pre- or self-diagnosis of PCOS and possibly provide the opportunity for users to be educated and informed effectively. Gibson-Helm et al [37,38] reported that many women experience a prolonged and frustrating diagnosis of PCOS because it requires multiple health care providers to

evaluate and diagnose PCOS, based on their study of a large community-based national sample. Our proposed work will resolve the concerns of delayed diagnosis and improve health care access for women.

Limitations and Future Works

The main limitation of this study is that the sample is from a very specific population in Kerala, India, which is a state along India's tropical Malabar coast. In previous studies, women experienced variations in PCOS symptoms depending on culture and ethnicity [38-40]. Another limitation is that the data set was collected from multiple hospitals where slightly different criteria might have been used to diagnose patients with PCOS, causing higher error rates in prediction. Our future work will include

collecting larger data sets from different cultures and ethnicities for comparison and improving our prediction models. This will enable us to provide more tailored prediction tools for users in different cultural and ethnic groups. In addition, there is abundant evidence that PCOS substantially contributes to women's anxiety disorders, depression, personality, and other psychological disorders [41,42]. A more in-depth study on the psychological effects of PCOS will be conducted in the future.

Conclusions

In summary, the proposed study offers great potential that our self-diagnostic prediction models for PCOS status can serve as a convenient and easy-to-use digital platform based on available health measures for both potential patients and clinical providers.

Conflicts of Interest

None declared.

References

1. Ting DS, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. Nat Med 2020 Apr 27;26(4):459-461 [FREE Full text] [doi: 10.1038/s41591-020-0824-5] [Medline: 32284618]
2. Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J, et al. Telehealth for global emergencies: implications for coronavirus disease 2019 (COVID-19). J Telemed Telecare 2020 Mar 20;26(5):309-313. [doi: 10.1177/1357633x20916567]
3. What to know about telehealth for Polycystic Ovary Syndrome (PCOS). Verywellhealth. 2021. URL: https://www.verywellhealth.com/telehealth-pcos-5114599 [accessed 2022-02-22]
4. Escobar-Morreale HF. Polycystic ovary syndrome: definition, aetiology, diagnosis and treatment. Nat Rev Endocrinol 2018 May 23;14(5):270-284. [doi: 10.1038/nrendo.2018.24] [Medline: 29569621]
5. Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). Hum Reprod 2004 Jan 01;19(1):41-47. [doi: 10.1093/humrep/deh098] [Medline: 14688154]
6. Teede H, Misso M, Costello M, Dokras A, Laven J, Moran L, International PCOS Network. Erratum. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. Hum Reprod 2019 Feb 01;34(2):388-1618 [FREE Full text] [doi: 10.1093/humrep/dey363] [Medline: 30521039]
7. Deeks AA, Gibson-Helm ME, Teede HJ. Anxiety and depression in polycystic ovary syndrome: a comprehensive investigation. Fertil Steril 2010 May 01;93(7):2421-2423. [doi: 10.1016/j.fertnstert.2009.09.018] [Medline: 20117778]
8. Toosy S, Sodi R, Pappachan JM. Lean polycystic ovary syndrome (PCOS): an evidence-based practical approach. J Diabetes Metab Disord 2018 Dec 13;17(2):277-285 [FREE Full text] [doi: 10.1007/s40200-018-0371-5] [Medline: 30918863]
9. Azziz R. Controversy in clinical endocrinology: diagnosis of polycystic ovarian syndrome: the Rotterdam criteria are premature. J Clin Endocrinol Metab 2006 Mar;91(3):781-785. [doi: 10.1210/jc.2005-2153] [Medline: 16418211]
10. Lujan ME, Chizen DR, Pierson RA. Diagnostic criteria for polycystic ovary syndrome: pitfalls and controversies. J Obstetrics Gynaecol Canada 2008 Aug;30(8):671-679. [doi: 10.1016/s1701-2163(16)32915-2]
11. Franks S. Controversy in clinical endocrinology: diagnosis of polycystic ovarian syndrome: in defense of the Rotterdam criteria. J Clin Endocrinol Metab 2006 Mar;91(3):786-789. [doi: 10.1210/jc.2005-2501] [Medline: 16418209]
12. Goyal M, Dawood A. Debates regarding lean patients with polycystic ovary syndrome: a narrative review. J Hum Reprod Sci 2017;10(3):154. [doi: 10.4103/jhrs.jhrs_77_17]
13. Moran LJ, Lombard CB, Lim S, Noakes M, Teede HJ. Polycystic ovary syndrome and weight management. Womens Health (Lond) 2010 Mar 01;6(2):271-283 [FREE Full text] [doi: 10.2217/whe.09.89] [Medline: 20187731]
14. Moran LJ, Norman RJ, Teede HJ. Metabolic risk in PCOS: phenotype and adiposity impact. Trends Endocrinol Metab 2015 Mar;26(3):136-143. [doi: 10.1016/j.tem.2014.12.003] [Medline: 25591984]
15. Alvarez-Blasco F, Botella-Carretero JI, San Millán JL, Escobar-Morreale HF. Prevalence and characteristics of the polycystic ovary syndrome in overweight and obese women. Arch Intern Med 2006 Oct 23;166(19):2081-2086. [doi: 10.1001/archinte.166.19.2081] [Medline: 17060537]
16. Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep learning-based multi-omics integration robustly predicts survival in liver cancer. Clin Cancer Res 2018 Mar 15;24(6):1248-1259 [FREE Full text] [doi: 10.1158/1078-0432.CCR-17-0853] [Medline: 28982688]

XSL•FO

RenderX

17.   Saber A, Sakr M, Abo-Seida OM, Keshk A, Chen H. A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique. IEEE Access 2021;9:71194-71209. [doi: 10.1109/access.2021.3079204]

18.   Álvarez JD, Matias-Guiu JA, Cabrera-Martín MN, Risco-Martín JL, Ayala JL. An application of machine learning with feature selection to improve diagnosis and classification of neurodegenerative disorders. BMC Bioinformatics 2019 Oct 11;20(1):491 [FREE Full text] [doi: 10.1186/s12859-019-3027-7] [Medline: 31601182]

19.   Maugeri A, Barchitta M, Basile G, Agodi A. Applying a hierarchical clustering on principal components approach to identify different patterns of the SARS-CoV-2 epidemic across Italian regions. Sci Rep 2021 Mar 29;11(1):7082 [FREE Full text] [doi: 10.1038/s41598-021-86703-3] [Medline: 33782519]

20.   Kumar S, Kumari K A, Mohapatra S, Naik B, Nayak J, Mishra M. CatBoost ensemble approach for diabetes risk prediction at early stages. In: Proceedings of the 2021 1st Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology(ODICON). 2021 Presented at: 2021 1st Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology(ODICON); Jan 8-9, 2021; Bhubaneswar, India. [doi: 10.1109/ODICON50556.2021.9428943]

21.   Aboueid S, Liu RH, Desta BN, Chaurasia A, Ebrahim S. The use of artificially intelligent self-diagnosing digital platforms by the general public: scoping review. JMIR Med Inform 2019 May 01;7(2):e13445 [FREE Full text] [doi: 10.2196/13445] [Medline: 31042151]

22.   Polycystic ovary syndrome (PCOS). Kaggle. URL: https://www.kaggle.com/prasoonkottarathil/polycystic-ovary-syndrome-pcos [accessed 2022-02-22]

23.   Dapas M, Lin FT, Nadkarni GN, Sisk R, Legro RS, Urbanek M, et al. Distinct subtypes of polycystic ovary syndrome with novel genetic associations: an unsupervised, phenotypic clustering analysis. PLoS Med 2020 Jun 23;17(6):e1003132 [FREE Full text] [doi: 10.1371/journal.pmed.1003132] [Medline: 32574161]

24.   Kar S. Anthropometric, clinical, and metabolic comparisons of the four Rotterdam PCOS phenotypes: a prospective study of PCOS women. J Hum Reprod Sci 2013 Jul;6(3):194-200 [FREE Full text] [doi: 10.4103/0974-1208.121422] [Medline: 24347934]

25.   Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci 2016 Apr 13;374(2065):20150202 [FREE Full text] [doi: 10.1098/rsta.2015.0202] [Medline: 26953178]

26.   James G, Witten D, Hastie T, Tibshirani R. Introduction. In: An Introduction to Statistical Learning. New York: Springer; 2013.

27.   Syakur MA, Khotimah BK, Rochman EM, Satoto BD. Integration K-means clustering method and elbow method for identification of the best customer profile cluster. In: The 2nd International Conference on Vocational Education and Electrical Engineering (ICVEE). 2017 Presented at: The 2nd International Conference on Vocational Education and Electrical Engineering (ICVEE); Nov 9, 2017; Surabaya, Indonesia. [doi: 10.1088/1757-899x/336/1/012017]

28.   CatBoost – A new game of machine learning. Affine. URL: https://affine.ai/catboost-a-new-game-of-machine-learning/ [accessed 2022-02-22]

29.   Lundberg S, Lee SI. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: Proceedings of the 31st International Conference on Neural Information Processing Systems; Dec 4 - 9, 2017; Long Beach California USA. [doi: 10.5555/3295222.3295230]

30.   Zabuliene L, Tutkuviene J. [Body composition and polycystic ovary syndrome]. Medicina (Kaunas) 2010;46(2):142-157 [FREE Full text] [Medline: 20440089]

31.   Sahmay S, Aydin Y, Oncul M, Senturk LM. Diagnosis of polycystic ovary syndrome: AMH in combination with clinical symptoms. J Assist Reprod Genet 2014 Feb 18;31(2):213-220 [FREE Full text] [doi: 10.1007/s10815-013-0149-0] [Medline: 24343036]

32.   Hsu M. Changes in the PCOS phenotype with age. Steroids 2013 Aug;78(8):761-766. [doi: 10.1016/j.steroids.2013.04.005] [Medline: 23624031]

33.   Farshchi H, Rane A, Love A, Kennedy RL. Diet and nutrition in polycystic ovary syndrome (PCOS): pointers for nutritional management. J Obstet Gynaecol 2007 Nov 02;27(8):762-773. [doi: 10.1080/01443610701667338] [Medline: 18097891]

34.   About adult BMI. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html#Why [accessed 2022-02-22]

35.   Kilit C. Heart rate variability in normal-weight patients with polycystic ovary syndrome. Anatol J Cardiol 2017. [doi: 10.14744/anatoljcardiol.2017.7473]

36.   Ji H, Woo H, Park Y, Hwang D, Lee J, Lee C, et al. Characteristics of heart rate variability in women with polycystic ovary syndrome: a retrospective cross-sectional study. Medicine (Baltimore) 2018 Sep;97(38):e12510 [FREE Full text] [doi: 10.1097/MD.0000000000012510] [Medline: 30235765]

37.   Gibson-Helm ME, Lucas IM, Boyle JA, Teede HJ. Women's experiences of polycystic ovary syndrome diagnosis. Fam Pract 2014 Oct 12;31(5):545-549. [doi: 10.1093/fampra/cmu028] [Medline: 24925927]

38.   Gibson-Helm M, Teede H, Dunaif A, Dokras A. Delayed diagnosis and a lack of information associated with dissatisfaction in women with polycystic ovary syndrome. J Clin Endocrinol Metab 2017 Feb 01;102(2):604-612 [FREE Full text] [doi: 10.1210/jc.2016-2963] [Medline: 27906550]

39. Engmann L, Jin S, Sun F, Legro RS, Polotsky AJ, Hansen KR, Reproductive Medicine Network. Racial and ethnic differences in the polycystic ovary syndrome metabolic phenotype. Am J Obstet Gynecol 2017 May;216(5):493.e1-493e13 [FREE Full text] [doi: 10.1016/j.ajog.2017.01.003] [Medline: 28104402]

40. Sendur SN, Yildiz BO. Influence of ethnicity on different aspects of polycystic ovary syndrome: a systematic review. Reprod Biomed Online 2021 Apr;42(4):799-818. [doi: 10.1016/j.rbmo.2020.12.006] [Medline: 33487557]

41. Sayyah-Melli M, Alizadeh M, Pourafkary N, Ouladsahebmadarek E, Jafari-Shobeiri M, Abbassi J, et al. Psychosocial factors associated with polycystic ovary syndrome: a case control study. J Caring Sci 2015 Sep 01;4(3):225-231 [FREE Full text] [doi: 10.15171/jcs.2015.023] [Medline: 26464839]

42. Månsson M, Holte J, Landin-Wilhelmsen K, Dahlgren E, Johansson A, Landén M. Women with polycystic ovary syndrome are often depressed or anxious--a case control study. Psychoneuroendocrinology 2008 Sep;33(8):1132-1138. [doi: 10.1016/j.psyneuen.2008.06.003] [Medline: 18672334]

Abbreviations

AMH: anti-Müllerian hormone
BP: blood pressure
PCA: Principal Component Analysis
PCOS: polycystic ovary syndrome
SHAP: Shapley Additive Explanations

Original Paper

Acceptance of the District Health Information System Version 2 Platform for Malaria Case-Based Surveillance By Health Care Workers in Botswana: Web-Based Survey

Kagiso Ndlovu[1*], MSc; Kabelo Leonard Mauco[2*], PhD; Mpho Keetile[3*], PhD; Khutsafalo Kadimo[4*], MSc; Refilwe Yvonne Senyatso[5*], MD; Davies Ntebela[5*], MSc; Buthugwashe Valela[5*], MSc; Clement Murambi[5], BSc

[1]Department of Computer Science, Faculty of Science, University of Botswana, Gaborone, Botswana

[2]Department of Health Information Management, Faculty of Health and Education, Botho University, Gaborone, Botswana

[3]Department of Population Studies, Faculty of Social Sciences, University of Botswana, Gaborone, Botswana

[4]Department of Library Services, University of Botswana, Gaborone, Botswana

[5]National Malaria Programme, Ministry of Health and Wellness, Gaborone, Botswana

[*]these authors contributed equally

Corresponding Author:
Kagiso Ndlovu, MSc
Department of Computer Science
Faculty of Science
University of Botswana
Private Bag UB 0022
Plot 4775 Notwane Rd
Gaborone, 00267
Botswana
Phone: 267 3555700 ext 5700
Fax: 267 3185097
Email: ndlovuk@ub.ac.bw

Abstract

Background: Similar to many low- and middle-income countries, Botswana has identified eHealth as a means of improving health care service provision and delivery. The National Malaria Programme (NMP) in Botswana has implemented the District Health Information System version 2 (DHIS2) to support timely malaria case reporting across its 27 health districts; however, the implementation of an eHealth system is never without challenges. Barriers to the implementation of eHealth innovations within health care settings may arise at the individual or organizational levels. As such, the evaluation of user perceptions of the technology is an important step that can inform its sustainable implementation. The DHIS2 was implemented without evaluating user perceptions beforehand; therefore, the Botswana Ministry of Health and Wellness was uncertain about the likelihood of acceptance and use of the platform.

Objective: We aimed to determine the acceptance of the DHIS2 platform by the NMP in Botswana to gauge whether adoption would be successful.

Methods: The study's design was informed by constructs of the technology acceptance model. A survey, with items assessed using a 7-point Likert scale, and focus group discussions were undertaken with DHIS2 core users from 27 health districts and NMP personnel at the Ministry of Health and Wellness. The web-based survey was administered from August 3, 2020 to September 30, 2020.

Results: Survey participants were core users (n=27). Focus group participants were NMP personnel (n=5). Overall, participants' survey responses (frequently occurring scores of 7) showed their confidence in the DHIS2 platform for case-based surveillance of malaria; however, participants also noted some organizational issues that could compromise user acceptance of the DHIS2 platform.

Conclusions: Participants' responses indicated their acceptance of the DHIS2 platform; however, the consideration of factors related to organizational readiness could further enhance successful acceptance, and consequently, successful adoption of the platform by the malaria program in Botswana.

XSL•FO
RenderX

Introduction

Botswana is among the countries that have made substantial progress, in the elimination continuum, in the fight against malaria. Between 2000 and 2015, the malaria incidence in Botswana fell dramatically, by 79%, from 0.136 in 2000 to 0.029 in 2015 [1]. Furthermore, mortality caused by malaria declined by 57%, from 1069 deaths in 2003 to 462 deaths in 2015 [2]. This reduction in malaria morbidity and mortality is mainly due to the government's implementation of malaria interventions [3,4] such as annual health care worker training on malaria case management to ensure accurate diagnosis and treatment of identified cases [5]. The transmission of malaria in Botswana is highly heterogeneous; transmission is mostly seasonal and unstable, occurring primarily between November and May during the rainy season [6] and primarily in the northern and eastern parts of the country [7]. However, malaria transmission in Botswana is very low compared to that in other sub-Saharan African countries [8]. Despite efforts toward malaria elimination, Botswana has not been successful in meeting its target of achieving zero indigenous cases [7]. Failure to meet the set target could be attributed to inadequate disease surveillance systems. This is because an adequate disease surveillance system will ensure that data collection, analysis, reporting, active case finding, and linkage to the response happens quickly to identify infections (symptomatic and asymptomatic), prevent ongoing transmission, and decrease the transmission efficiency of vectors. Similar to many low- and middle-income countries, Botswana has identified eHealth (ie, "the use of Information and Communications Technologies (ICT) for health" [9]) as a means of improving health care service provision and delivery, and other low- and middle-income countries have implemented eHealth systems for case-based surveillance of malaria [10,11]. Such trends continue to be spurred by the recent global awareness and application of eHealth technologies toward combating the current COVID-19 pandemic [12].

In Botswana, an eHealth system—the District Health Information System version 2 (DHIS2)—has been identified to improve malaria surveillance across the 27 health districts. The decision to consider DHIS2 was based on its documented benefits [13-15]. In spite of the availability of the DHIS2 platform for case-based surveillance of malaria, paper-based case notification forms continue to be utilized. Prior to the DHIS2, the National Malaria Programme (NMP) used a spreadsheet (Excel, version 2013; Microsoft Inc) for data capture, analysis, and reporting. At the district level, either monitoring and evaluation officers, a malaria focal person (the person who is responsible for management of the NMP in the district), or a community health nurse compiled information on malaria infections (patient name, age, date of diagnosis, physical address, the treatment offered, and method of detection) in the spreadsheet, based on presentations and diagnoses in the preceding week, which was then emailed to the NMP at the Ministry of Health and Wellness headquarters. In addition, a designated health care worker at each facility could complete a case notification paper form (in duplicate), which was sent to the district headquarters and to the NMP. The monitoring and evaluation officer compiled data from all 27 health districts in a spreadsheet arranged by year, month, district, and week.

Notwithstanding the potential benefits of the DHIS2 platform toward improving case-based surveillance of malaria in Botswana, the implementation of any eHealth system is never without challenges [16]. According to Ross et al [17], barriers to implementation of eHealth innovations within health care settings may arise at the individual, organization, or wider levels of the health care system. Acceptance by users (health care workers) has been documented as one of the key factors for successful implementation of an eHealth system [18]; therefore, evaluating user perceptions of the technology is an important step. The technology acceptance model features prominently among key theoretical approaches used to understand people's intentions to accept various forms of information technology [19,20]. Many studies [20-25] that focus on explaining end user acceptance and predicting successful adoption of eHealth systems by health care organizations use the technology acceptance model—that an individual's (1) acceptance of (intention to engage with) a technology depends on (2) perceived usefulness and (3) perceived ease of use [19-21]—as a basis. This model contends that a strong relationship exists between one's intention to use technology and their actual usage behavior [19,20], and perceived usefulness is characterized by an individual's belief that engaging a technology improves their job performance, while perceived ease of use refers to their belief that using technology requires minimal effort [19,20].

The DHIS2 platform for case-based surveillance of malaria in Botswana was implemented without evaluating user acceptance. Evaluation of user acceptance is necessary to inform sustainable implementation of technology. We aimed to evaluate user acceptance of the DHIS2 platform.

Methods

Study Design

Survey items and the focus group discussion guide were informed by constructs (acceptance of the technology, perceived usefulness, and perceived ease of use) of the technology acceptance model and were used to gain insights from users across the 27 health care districts in Botswana.

We developed a survey and focus group discussion guide, both in English. The drafts were first reviewed by a health care worker at the Ministry of Health and Wellness and by a medical librarian at the University of Botswana, after which, both tools were refined. The revised tools were then tested by 2 other

health care workers, whose input resulted in further enhancement of the tools through improved branching logic. The final survey was administered using REDCap (Vanderbilt University [23]) forms from August 3, 2020 to September 30, 2020. The final version of the focus group questions was utilized to guide focus group discussions.

For both the survey and focus group, we used purposive sampling. The DHIS2 was put into place across the 27 health districts in Botswana, with each district having at least one core user (a health care worker whose main role involves dedicated interaction with the DHIS2 platform for the NMP). The sample consisted 1 core user from each health districts and NMP personnel at the Ministry of Health and Wellness headquarters who were part of a unit dedicated to the use of DHIS2 platform for the NMP in Botswana (the NMP coordinator, a health informatics officer, a software developer, and 2 monitoring and evaluation officers).

Participants from the 27 health districts were recruited by the NMP coordinator, and those who consented were sent a link to complete the survey. Only one author was responsible for accessing and managing the web-based database through the use of username and password. The survey consisted of questions about the type of facility where the respondent was based and the name of the district and of statements that assessed users' acceptance of the DHIS2 platform, perceived usefulness, and perceived ease of use, using a 7-point Likert scale. There were 30 close-ended statements and 4 open-ended questions.

Focus group discussions, 1 session which lasted for 1 hour, involved the 5 NMP personnel based at the Ministry of Health and Wellness, who were invited by email and text messaging to a videoconference (Google Meet, Google LLC; which was chosen because it was the only platform to which all participants had access). The purpose of the focus group discussion was explained to participants, after which, they were asked to provide consent to participate in the study. The focus group discussion was recorded (with participant permission) and later transcribed verbatim.

Descriptive statistics (mean, standard deviation, median, and mode) were calculated (Excel, version 2013, Microsoft Inc) for quantitative data (close-ended statement ratings). Qualitative data (from open-ended questions and focus group discussion) were analyzed using thematic analysis (NVivo, version 11; QSR International). Thematic analysis of participants' responses was conducted to determine factors affecting the sustainable implementation of the DHIS2 platform for the case-based surveillance program. We used a deductive coding process [22]. A coding frame and a predefined list of descriptive codes were developed by 1 author and then discussed by all authors, which yielded 20 codes ("eHealth," "mHealth," "ICT," "user acceptance," "usefulness," "ease of use," "user interface," "capacity development," "internet connectivity," "mobile devices," "electronic medical record," "electronic health record," "security," "privacy," "confidentiality," "interoperability," "integration," "data management," "data analysis," and "data reporting"). Transcriptions of the discussions were systematically and iteratively searched (in 2 cycles) for elements relevant to the 20 codes. Participants' responses from focus group discussions, as well as open-ended responses from the survey, were matched to corresponding categories in the coding frame. This process was done iteratively to refine the alignment and identify high-level themes.

Ethics

The study was approved by the Office of Research and Development of the University of Botswana (UBR/RES/IRB/BIO/224) and the Ministry of Health and Wellness (HPDME: 13/18/1). Participants provided informed consent. During data collection, we presented and explained the consent form to seek permission from potential participants. The consent forms also clearly explained the purpose of the study and provided assurance that data would be kept safe and deidentified. Participants were informed of their right to refuse to participate or withdraw from the study at any time.

Results

The survey response rate was 89% (24/27). From a total of 27 DHIS2 core users (community health nurses: n=14; malaria focal person: n=9; health information and communication technology personnel: n=4), 24 responded (community health nurses n=13; malaria focal persons: n=7; health information and communication technology personnel: n=4). Of those who responded, 15 were from the district health management team, 3 were from public hospitals, and 6 were from public clinic facilities. Core users who did not respond were from public hospitals (n=2) and a public clinic (n=1). Participants most frequently responded that they agreed (4 items: mode 6) or strongly agreed (16 items: mode 7) with survey statements, which were aligned with technology acceptance model constructs (Table 1).

All 5 NMP personnel participated in the focus group and highlighted possible factors that could affect the sustainable implementation of the DHIS2 platform for case-based surveillance of malaria (Table 2). Ultimately, 5 themes were identified (governance, infrastructure, capacity building, data security, and usability).

Table 1. Acceptance of the District Health Information System version 2 based on the technology acceptance model.

Constructs[a]	Mean (SD)	Median	Mode
Perceived ease of use			
Overall, I am satisfied with how easy it is to use DHIS2[b]	5.2 (1.9)	6	7
It was simple to use DHIS2	5.4 (2.0)	6	7
It is easy to find the information I needed	5.2 (1.7)	5	7
It was easy to learn to use DHIS2	5.3 (2.2)	7	7
I feel comfortable using DHIS2	5.7 (2.1)	7	7
The information (such as online help, on-screen messages, and other documentation) provided with DHIS2 is clear	5.1 (1.7)	5	7
The interface of DHIS2 is pleasant	5.7 (1.5)	6	7
DHIS2 gives error messages that clearly tell me how to fix problems	3.8 (2.2)	3	3
Whenever I make a mistake using DHIS2, I recover easily and quickly	4.6 (1.9)	4	3
The organization of information on the system screens is clear	5.8 (1.3)	6	6
The information provided for the system is easy to understand	5.3 (1.9)	6	7
Perceived usefulness			
I am able to complete my work quickly using DHIS2	5.1 (2.2)	6	7
I can effectively complete my work using DHIS2	5.5 (1.9)	7	7
I believe I became productive quickly using DHIS2	5.4 (1.9)	6	7
The information is effective in helping me complete the tasks and scenarios	5.4 (1.8)	6	7
DHIS2 has all the functions and capabilities I expect it to have	5.4 (1.5)	6	6
Overall, DHIS2 improved data management activities for the malaria program	5.6 (2.1)	7	7
Overall, DHIS2 improved data analysis of malaria cases	5.6 (1.6)	6	7
DHIS2 contributed to timely decision-making processes	5.9 (1.6)	6	7
Overall, DHIS2 improved timely reporting of malaria cases	5.8 (1.8)	6	7
Acceptance of technology			
I like using the interface of DHIS2	5.1 (1.9)	6	6
Overall, I am satisfied with the DHIS2 system	5.1 (1.9)	6	6

[a]Responses were in the form of a Likert scale from 1 (strongly disagree) to 7 (strongly agree).

[b]DHIS2: District Health Information System version 2.

Table 2. Thematic presentation of factors affecting the sustainable implementation of the District Health Information System version 2 (DHIS2) platform for case-based surveillance of malaria.

Theme	Example quotes
Governance	"It adds on the already heavy workload, there should be people specific to do malaria data entry"
	"It brought about migration from manual system to electronic platform given that COVID-19 necessitates the use of technology."
	"Have government employees being part of the system programming and maintenance/being part of the project or have ownership to the system"
	"Majority of health care workers should be shown the importance of using the health systems"
	"Cascade mentoring of DHIS utilization in facilities as facility based training"
	"Every time an update is introduced on the paper data collection tools (guidelines), the same should also be updated on the DHI2 system to make sure that the paper forms and the system are in sync at all times".
Infrastructure	"Frequent network downtimes resulting in delays in reporting when the system is down"
	"It is not linked with other applications that the government has created to formulate reports and daily activities"
	"Data capture even when there is no connectivity and later sync when the internet connection is regained"
	"More gadgets should be availed for each facility and errors or SIM card malfunctions should be urgently addressed"
	"System linkage/interoperability with other existing systems"
	"Making sure internet is reliable at facilities"
Capacity building	"Requires more practice for better or improved competencies"
	"Capacity building to be strengthened for the local IT Officers"
	"Training should have frequent support visits"
	"Cascade mentorship programmes of DHIS utilization in facilities as facility-based training"
	"All health care workers should be trained and shown the importance of using the health information systems"
	"Topics covered to be taken step by step or explained in full with more practicals"
	"Training content coverage should be more than 80% practicals"
	"There should have been site visit immediately after training to assess situations on the ground"
Data security	"Use of authentication passwords and encryption algorithms for data security and privacy"
	"The link is not trustworthy"
Usability	"It picks everything you enter whether good or bad"
	"No free choice to select the period you want to view the data, all are fixed to daily, weekly, monthly, etc"
	"It's confusing sometimes as one can enter patients twice but not knowing"
	"Clients are attached to the diagnosing facility rather than the origin of infection"
	"If an error is made during entering of cases and you want to correct it, there isn't an option to edit especially on the date of diagnosis and notification"
	"Data management greatly improved for the intervention of positive malaria cases"
	"DHIS2 is easy to use as it simplifies data management and analysis"
	"It is pleasant, easier and quick to use, you can retrieve your information if you need to utilize it for something"
	"Timely data reporting"
	"Completeness of reports is achieved"
	"Easy analysis of data, using maps and graphs for visualisations"
	"Easy to formulate reports if used adequately and effectively"
	"It covered most aspects of malaria case-based surveillance"
	"DHIS2 has simplified malaria surveillance and indeed it is a very valuable tool"
	"Gives a good situational analysis of what is happening on the ground"
	"DHIS2 enables foursight mapping, by using the GPS coordinates to locate malaria case outbreaks or hotspots."
	"Everything is now at our fingertips, we have data in the form of tables, charts and reports generated from the dashboard than prior to the DHIS2 where we used to manually manage the data using excel charts."
	"We easily make decisions based on the data from the system"

Discussion

General

Overall, the study showed participants' satisfaction with the DHIS2 platform, with the majority of responses indicating agreement (20/22; 4 items: mode 6; 16 items: mode 7) with positive statements based on technology acceptance model constructs (Table 1). Key themes—*governance*, *infrastructure*, *capacity building*, *data security*, and *usability*—related to an

organization's role in influencing technology acceptance were identified (Table 2).

The technology acceptance model constructs of perceived ease of use and perceived usefulness have been previously documented as positive influences toward technology adoption [24]. Moreover, *perceived ease of use* has been defined as the extent to which a person believes that using a technology will be free of effort [24], and *perceived usefulness* is the belief that one's utilization of information technologies will enhance one's work performance [25]. For only 2 statements (addressing system usability) of the 22 survey items did the most common response by participants indicate that they somewhat disagreed (mode 3). Based on our findings, successful acceptance and adoption of the DHIS2 platform by the malaria program in Botswana could be assumed; however, NMP personnel highlighted some factors that could negatively influence acceptance of the DHIS2 for the malaria program. In an organization, factors that may affect sustainable implementation of the DHIS2 platform for case-based surveillance of malaria include the state of governance, the state of relevant infrastructure, the presence or absence of capacity building initiatives aimed at empowering potential users of the system, data security measures, and usability of the system (Table 2). As such, the application of project management practices by an implementing organization may be critical in ensuring sustainable implementation of the DHIS2 platform. Successful project management practices will ensure the fitness of a project for its political context (ie, in terms of organizational strategy, managership, and stakeholder management) [26]. Mlekus et al [27] highlighted that any organization planning for the successful acceptance and adoption of new technology should consider issues that fulfill user experience related to output quality, perspicuity, dependability, and novelty. Change management has been documented as one strategy that organizations could use to raise acceptance of new technology in a workplace and hence improve chances of successful adoption of the technology [28]. Change management is about supporting people through a process of change, and successful implementation of change is achieved when the systems, processes, tools, and technology of the change initiative are embedded in the new way in which health care providers do their work [29]. Ingebrigsten et al [30] identified 7 leadership behaviors that were associated with successful outcomes in Health Information Technology adoption: (1) communicating clearly about visions and goals, (2) providing support, (3) establishing a governance structure, (4) establishing training, (5) identifying and appointing champions, (6) addressing work process change, and (7) following up. Such leadership maybe necessary in addressing the issues highlighted in Table 2. Consequently, a lack of top management support and technology implementation strategy may play a negative role in influencing information and communication technology acceptance and adoption in any organization. Hu et al [31] noted that the ultimate success of an eHealth system in an adopting organization requires adequate attention to both technological and managerial issues. Therefore, for successful and sustainable implementation of an eHealth system, the health care organization should ensure availability of requisite resources

and processes. In fact, failures of eHealth system implementation have been associated with a lack of eHealth readiness [32,33].

The technology acceptance model is one of the most widely used theoretical frameworks for predicting individuals' likelihood to accept and adopt new technology [34]. The model is based on the assumption that when users perceive that a type of technology is useful and easy to use, they will be willing to use it [35]. However, some of our findings suggest the need for a model that considers the role of organizational readiness, the extent to which an institutional setting and culture support and promotes awareness, implementation, and use of eHealth [36], in influencing technology acceptance and hence the adoption of eHealth systems.

Guidance

In line with key study findings, the following acceptance-related issues and associated mitigation strategies are proffered: Perception of usability can influence acceptance of an eHealth system. As such, the usability of the current DHIS2 platform for case-based surveillance of malaria should be enhanced as it could deter users from interacting with the platform. One focus group participant reported usability concerns about the DHIS2,

> *It's confusing sometimes as one can enter patients twice but not knowing.*

Another participant reported that

> *If an error is made during entering of cases and you want to correct it, there isn't an option to edit especially on the date of diagnosis and notification.*

Previous studies [37-40] have suggested enhancing system usability with user-friendly interfaces, real-time feedback mechanisms, and decision-support capabilities. It is, therefore, important to involve all key stakeholders, from design to implementation, by using a user-centered design approach [41]. User-centered design is also considered to be an important contributor to system usefulness and usability [41]. This is of importance since user-centered design creates a sense of ownership among end users who participated in the design and development process, thereby increasing their acceptance of the system

Digital health literacy is also a factor in technology acceptance [42]. Study participants highlighted the need for frequent training on the DHIS2 platform, with one stating

> *Training should have frequent support visits*

while another suggested the DHIS2 platform

> *...requires more practice for better or improved competencies.*

This could be achieved through continuous health human resource capacity-building programs as technology and user needs to evolve.

A potential data security consideration, the

> *...use of authentication passwords and encryption algorithms for data security and privacy...*

was identified in this study, as well as, one security risk, that is,

The link is not trustworthy.

These issues require comprehensive well-documented regulatory approaches to facilitate protection of the highly sensitive clinical data. Given the limited documented data security and privacy best practices for open-source solutions, implementation of the DHIS2 platform in Botswana should be guided by and aligned to key national policy documents such as the Data Protection Act [43] or equivalent policy documents in other settings.

Focus group participants indicated that the DHIS2 could be enhanced by establishing

> *...system linkage/interoperability with other existing systems.*

This followed a previous finding [44] highlighting that the DHIS2 is not linked with other applications that the government has created to formulate reports and daily activities. Interoperability of eHealth systems in Botswana and similar low- and middle-income countries should be strengthened through increased adaptation of universally available software, services, and content, such as the DHIS2 platform, as these are already endorsed by the World Health Organization to be interoperable [9]. Furthermore, the DHIS2 supports a web application programming interface that allows for integration with other databases and supports the development of an "Integrated Information Portal [45]." Lastly, it is important to view interoperability of eHealth systems as an ongoing process that can be improved over time.

Conclusions

Health care environments in most low- and middle-income countries are generally data rich but information poor. In Botswana, the health care sector is compounded with vast amounts of clinical data which are seldom used for decision-making purposes. The DHIS2 is a platform that can improve data capture, analysis, and timely reporting, which will facilitate transitioning data to information. Evaluation of the DHIS2 platform for case-based surveillance of malaria by core users showed that it was accepted. There is, however, room to improve the usability of the current DHIS2 platform, through end user involvement and buy in, to ensure sustainable health human resource capacity building, to address security privacy and confidentiality issues, and to address interoperability, by considering guidance from the World Health Organization [9]. Our findings can be used to inform policy makers and health informatics leaders in Botswana and in similar low- and middle-income countries to successfully plan and implement effective eHealth platforms.

References

1. Motlaleng M, Edwards J, Namboze J, Butt W, Moakofhi K, Obopile M, et al. Driving towards malaria elimination in Botswana by 2018: progress on case-based surveillance, 2013-2014. Public Health Action 2018 Apr 25;8(Suppl 1):S24-S28 [FREE Full text] [doi: 10.5588/pha.17.0019] [Medline: 29713590]
2. World malaria report 2015. World Health Organization. 2015. URL: https://apps.who.int/iris/handle/10665/200018 [accessed 2022-02-21]
3. Botswana conducts a comprehensive malaria performance programme review: towards elimination. World Health Organization, Botswana. 2017 Oct 9. URL: https://tinyurl.com/2p9ed8ax [accessed 2022-02-20]
4. Simon C, Moakofhi K, Mosweunyane T, Jibril H, Nkomo B, Motlaleng M, et al. Malaria control in Botswana, 2008-2012: the path towards elimination. Malar J 2013 Dec 20;12:458-458 [FREE Full text] [doi: 10.1186/1475-2875-12-458] [Medline: 24359262]
5. Chihanga S, Haque U, Chanda E, Mosweunyane T, Moakofhi K, Jibril H, et al. Malaria elimination in Botswana, 2012-2014: achievements and challenges. Parasit Vectors 2016 Feb 24;9:99 [FREE Full text] [doi: 10.1186/s13071-016-1382-z] [Medline: 26911433]
6. Moakofhi K, Edwards JK, Motlaleng M, Namboze J, Butt W, Obopile M, et al. Advances in malaria elimination in Botswana: a dramatic shift to parasitological diagnosis, 2008-2014. Public Health Action 2018 Apr 25;8(Suppl 1):S34-S38 [FREE Full text] [doi: 10.5588/pha.17.0017] [Medline: 29713592]
7. Updated malaria strategic plan 2010–2018: towards malaria elimination. Ministry of Health and Wellness. 2014. URL: https://endmalaria.org/sites/default/files/botswa2010-2015.pdf [accessed 2022-02-26]

8.   World malaria report 2014. World Health Organization. 2014. URL: http://www.who.int/malaria/publications/ [accessed 2022-02-21]

9.   Digital health. World Health Organization. URL: https://www.who.int/health-topics/digital-health#tab=tab_1 [accessed 2022-02-22]

10.   Prue CS, Shannon KL, Khyang J, Edwards LJ, Ahmed S, Ram M, et al. Mobile phones improve case detection and management of malaria in rural Bangladesh. Malar J 2013;12:48 [FREE Full text] [doi: 10.1186/1475-2875-12-48] [Medline: 23374585]

11.   Ma S, Lawpoolsri S, Soonthornworasiri N, Khamsiriwatchara A, Jandee K, Taweeseneepitch K, et al. Effectiveness of implementation of electronic malaria information system as the national malaria surveillance system in Thailand. JMIR Public Health Surveill 2016 May 06;2(1):e20 [FREE Full text] [doi: 10.2196/publichealth.5347] [Medline: 27227156]

12.   Brørs G, Norman CD, Norekvål TM. Accelerated importance of eHealth literacy in the COVID-19 outbreak and beyond. Eur J Cardiovasc Nurs 2020 Aug 15;19(6):458-461 [FREE Full text] [doi: 10.1177/1474515120941307] [Medline: 32667217]

13.   Sahay S, Rashidian A, Doctor H. Challenges and opportunities of using DHIS2 to strengthen health information systems in the Eastern Mediterranean Region: A regional approach. E J Info Sys Dev Countries 2019 Jun 21;86(1):e12108. [doi: 10.1002/isd2.12108]

14.   Dehnavieh R, Haghdoost A, Khosravi A, Hoseinabadi F, Rahimi H, Poursheikhali A, et al. The district health information system (DHIS2): a literature review and meta-synthesis of its strengths and operational challenges based on the experiences of 11 countries. Health Inf Manag 2019 May;48(2):62-75. [doi: 10.1177/1833358318777713] [Medline: 29898604]

15.   Sahay J, Sahay S, Mukherjee A. Sustainability qualifiers of health management information systems implementation: case study of DHIS2 in India. In: Bandi RK, Ranjini CR, Klein S, Madon S, Monteiro E, editors. The Future of Digital Work: The Challenge of Inequality. Cham: Springer; Dec 2020:194-206.

16.   Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and facilitators to the implementation of ehealth services: systematic literature analysis. J Med Internet Res 2019 Nov 22;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]

17.   Ross LE, Vigod S, Wishart J, Waese M, Spence JD, Oliver J, et al. Barriers and facilitators to primary care for people with mental health and/or substance use issues: a qualitative study. BMC Fam Pract 2015 Oct 13;16:135 [FREE Full text] [doi: 10.1186/s12875-015-0353-3] [Medline: 26463083]

18.   Cilliers L, Flowerday S. User acceptance of telemedicine by health care workers a case of the Eastern cape province, South Africa. Electron J Inf Syst Dev Ctries 2017 Dec 05;65(1):1-10. [doi: 10.1002/j.1681-4835.2014.tb00467.x]

19.   Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]

20.   Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A systematic review of the technology acceptance model in health informatics. Appl Clin Inform 2018 Jul 15;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]

21.   Kamal SA, Shafiq M, Kakria P. Investigating acceptance of telemedicine services through an extended technology acceptance model (TAM). Technol Soc 2020 Feb;60:101212. [doi: 10.1016/j.techsoc.2019.101212]

22.   Skjott Linneberg M, Korsgaard S. Coding qualitative data: a synthesis guiding the novice. Qual Res J 2019 Jul 24;19(3):259-270. [doi: 10.1108/qrj-12-2018-0012]

23.   Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul 30;95(9):103208-103893 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]

24.   Brown ITJ. Individual and technological factors affecting perceived ease of use of web-based learning technologies in a developing country. Electron J Inf Syst Dev Ctries 2017 Dec 05;9(1):1-15. [doi: 10.1002/j.1681-4835.2002.tb00055.x]

25.   Baki R, Birgoren B, Aktepe A. A meta analysis of factors affecting perceived usefulness and perceived ease of use in the adoption of e-learning systems. Turkish Online J Distance Educ 2018:4-42. [doi: 10.17718/tojde.471649]

26.   Warth LL, Dyb K. eHealth initiatives; the relationship between project work and institutional practice. BMC Health Serv Res 2019 Jul 24;19(1):520 [FREE Full text] [doi: 10.1186/s12913-019-4346-0] [Medline: 31340819]

27.   Mlekus L, Bentler D, Paruzel A, Kato-Beiderwieden A, Maier G. How to raise technology acceptance: user experience characteristics as technology-inherent determinants. Gr Interakt Org 2020 Aug 06;51(3):273-283. [doi: 10.1007/s11612-020-00529-7]

28.   Aziz F, Aizuddin A, Rami M, Mahadi N, Rami AM, Razali F. The influence of leadership style towards technology acceptance in organization. Int J Adv Sci Technol 2020:218-225 [FREE Full text]

29.   Canada Health Infoway. A framework and toolkit for managing ehealth change: people and processes. Registered Nurses' Association of Ontario. 2013. URL: https://bpgordersettoolkit.rnao.ca/sites/default/files/CHI_ChangeMgmtGuide_ENG.pdf [accessed 2022-02-20]

30.   Ingebrigtsen T, Georgiou A, Clay-Williams R, Magrabi F, Hordern A, Prgomet M, et al. The impact of clinical leadership on health information technology adoption: systematic review. Int J Med Inform 2014 Jun;83(6):393-405. [doi: 10.1016/j.ijmedinf.2014.02.005] [Medline: 24656180]

31. Hu P, Chau P, Sheng O. Adoption of telemedicine technology by health care organizations: an exploratory study. J Organ Comput Electron Commer 2009 Nov 18;12(3):197-221. [doi: 10.1207/S15327744JOCE1203_01]

32. Mauco KL, Scott RE, Mars M. Validation of an e-health readiness assessment framework for developing countries. BMC Health Serv Res 2020 Jun 23;20(1):575 [FREE Full text] [doi: 10.1186/s12913-020-05448-3] [Medline: 32576174]

33. Khoja S, Scott RE, Casebeer AL, Mohsin M, Ishaq A, Gilani S. e-Health readiness assessment tools for healthcare institutions in developing countries. Telemed J E Health 2007 Aug;13(4):425-431. [doi: 10.1089/tmj.2006.0064] [Medline: 17848110]

34. Klaic M, Galea MP. Using the technology acceptance model to identify factors that predict likelihood to adopt tele-neurorehabilitation. Front Neurol 2020 Dec 2;11:580832 [FREE Full text] [doi: 10.3389/fneur.2020.580832] [Medline: 33343488]

35. Ajibade P. Technology acceptance model limitations and criticisms: exploring the practical applications and use in technology-related studies, mixed-method, and qualitative researches. Libr Philos Pract 2018 [FREE Full text]

36. Mauco KL, Scott RE, Mars M. Development of an ehealth readiness assessment framework for Botswana and other developing countries: interview study. JMIR Med Inform 2019 Aug 22;7(3):e12949 [FREE Full text] [doi: 10.2196/12949] [Medline: 31441429]

37. El-Sappagh S, Ali F, Hendawi A, Jang J, Kwak K. A mobile health monitoring-and-treatment system based on integration of the SSN sensor ontology and the HL7 FHIR standard. BMC Med Inform Decis Mak 2019 May 10;19(1):97-36 [FREE Full text] [doi: 10.1186/s12911-019-0806-z] [Medline: 31077222]

38. Rubio Ó, Trigo JD, Alesanco Á, Serrano L, García J. Analysis of ISO/IEEE 11073 built-in security and its potential IHE-based extensibility. J Biomed Inform 2016 Apr;60:270-285 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.006] [Medline: 26883877]

39. Ndlovu K, Mars M, Scott RE. Interoperability frameworks linking mHealth applications to electronic record systems. BMC Health Serv Res 2021 May 13;21(1):459-410 [FREE Full text] [doi: 10.1186/s12913-021-06473-6] [Medline: 33985495]

40. Memon M, Wagner S, Pedersen C, Beevi F, Hansen F. Ambient assisted living healthcare frameworks, platforms, standards, and quality attributes. Sensors (Basel) 2014 Mar 04;14(3):4312-4341 [FREE Full text] [doi: 10.3390/s140304312] [Medline: 24599192]

41. Mao J, Vredenburg K, Smith PW, Carey T. The state of user-centered design practice. Commun ACM 2005 Mar 01;48(3):105-109. [doi: 10.1145/1047671.1047677]

42. Kuek A, Hakkennes S. Healthcare staff digital literacy levels and their attitudes towards information systems. Health Informatics J 2020 Mar 15;26(1):592-612 [FREE Full text] [doi: 10.1177/1460458219839613] [Medline: 30983476]

43. Data protection act, 2018. Botswana Communications Regulatory Authority. URL: https://www.bocra.org.bw/sites/default/files/documents/DataProtectionAct.pdf [accessed 2022-02-21]

44. Ndlovu K, Scott RE, Mars M. Interoperability opportunities and challenges in linking mhealth applications and erecord systems: Botswana as an exemplar. BMC Med Inform Decis Mak 2021 Aug 21;21(1):246 [FREE Full text] [doi: 10.1186/s12911-021-01606-7] [Medline: 34419020]

45. Garg R, Garg A. District health information system (DHIS2) software in India. Adv Comput Sci Inf Technol 2015;2(10):39-42 [FREE Full text]

Abbreviations

DHIS2: District Health Information System version 2
NMP: National Malaria Programme

XSL•FO
RenderX

Patients' Experiences of Using an eHealth Pain Management Intervention Combined With Psychomotor Physiotherapy: Qualitative Study

Anne-Grethe Eiken[1], MSc; Dag Ø Nordanger[1], PhD; Lise Solberg Nes[2,3,4], PhD; Cecilie Varsi[2,5], PhD

[1]Department of Nursing and Health Promotion, Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway

[2]Department of Digital Health Research, Division of Medicine, Oslo University Hospital, Oslo, Norway

[3]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway

[4]Department of Psychiatry and Psychology, College of Medicine and Science, Mayo Clinic, Rochester, MN, United States

[5]Faculty of Health and Social Sciences, University of South-Eastern Norway, Drammen, Norway

Corresponding Author:
Cecilie Varsi, PhD
Department of Digital Health Research
Division of Medicine
Oslo University Hospital
Trondheimsveien 235
Oslo, 0424
Norway
Phone: 47 90856339
Email: cecilie.varsi@usn.no

Abstract

Background: Chronic pain is a major health challenge to those affected. Blended care with psychomotor physiotherapy (PMP) combined with eHealth self-management might be beneficial.

Objective: This study aims to explore how patients with chronic pain experience the combination of PMP and the use of EPIO, an eHealth self-management intervention for chronic pain.

Methods: Individual semistructured interviews were conducted with 5 adult patients with chronic pain (ie, participants) who used EPIO in combination with PMP over a period of 10 to 15 weeks. Interviews explored participants' experiences using this treatment combination in relation to their pain and analyzed their experiences using systematic text condensation.

Results: Participants described having benefited from using EPIO in combination with PMP in terms of increased awareness of bodily signals and how pain was related to stress and activity. They also described changes in the relationship to themselves in terms of increased self-acceptance, self-assertion, and hope and their relationship to their pain in terms of seeing pain as less harmful and engaging in more active coping strategies.

Conclusions: Results indicate that a blended care approach combining eHealth self-management interventions such as EPIO with PMP may be of value to patients living with chronic pain.

Trial Registration: ClinicalTrials.gov NCT03705104; https://clinicaltrials.gov/ct2/show/NCT03705104

Introduction

Background

As much as 30% of the Norwegian population reports living with chronic pain [1-3], defined as pain lasting for more than 3 months [4]. Pain is one of the most common reasons for why people seek health care [5,6], sick leave, and disability pensions [7-10], and is also associated with increased mortality [11].

Chronic pain is complex, with physical, social, and psychological aspects interacting and impacting individual experience and often carrying major consequences for those

affected and their social network [1,12]. Commonly reported pain-related challenges include sleep disturbances [13,14], anxiety and depression [13,15,16], and impairments of daily activities [13,17] and quality of life [13,18,19].

In multidisciplinary chronic pain management, the focus is generally on maximizing function and quality of life and may include combinations of medication treatment, physical therapy, and additional self-management approaches such as working with one's own thoughts, perceptions, attitudes, emotions, and activities [20,21].

Self-management focuses on supporting people to actively manage their own health and live as well as possible with their chronic condition [22]. This includes problem solving relevant for dealing with the consequences of the chronic condition and making necessary changes in daily life to provide improved quality of life [22]. In recent years, active self-management has been increasingly considered an important factor for pain management. There is consistent evidence that the use of active pain self-management strategies, such as maintaining activities despite pain, causes less pain-related disability than those using more passive approaches, such as trusting others [23], and recent clinical guidelines also recommend including self-management interventions in chronic pain treatment [20].

Self-management for people living with chronic pain (ie, actively managing their own health) can be affected by biomedical factors (eg, bodily symptoms) and the psychosocial aspects of pain (eg, emotional distress, self-efficacy, motivation for change, acceptance, worst-case thinking, fear of pain, helplessness, and barriers to pain management) [24-26].

Many patients living with chronic pain are offered psychomotor physiotherapy (PMP) [27]. PMP focuses on changing and facilitating insight and awareness for the patient to recognize and understand the connections between their daily life and their physical condition [27]. On the basis of the notion that the body and psyche are deeply interconnected, PMP treatment aims to raise awareness and change the states of tension in the body, with the intention of increasing the patient's familiarity and contact with his or her own body.

Interventions with cognitive behavioral therapy (CBT) [28] and acceptance and commitment therapy (ACT) [29] may also support self-management of chronic pain. CBT centers around the impact and relationships among thoughts, feelings, and behaviors and involves techniques to challenge and change thoughts and behaviors [28]. CBT approaches to chronic pain management have been associated with improved self-management; reductions in pain, disability, and emotional distress [30,31]; and improvements in mood, coping, and social functioning [30]. ACT focuses more on acceptance of a situation or setting, commitment to change, and attention to helpful strategies, with the goal of increasing psychological flexibility (eg, ability to adapt to situations and demands) [29,32] and has been associated with improvements in pain acceptance or psychological flexibility, pain intensity, disability or physical functioning, depressive symptoms, and quality of life [33,34].

In recent years, there has been an increasing interest in, and development of, eHealth CBT- and ACT-based interventions

for pain management. The delivery of such interventions is not dependent on access to professionals with pain management competence, treatment waiting lists, or travel distances and may be less associated with stigma than more traditional forms of psychological treatment [13,35]. Although some eHealth interventions have been associated with decreased pain intensity and improved function and quality of life [35,36], the field of eHealth pain management is still at an early stage and most available eHealth programs so far lack a theoretical foundation and have not been subjected to research and rigorous efficacy testing [36,37].

Studies have shown that blended care (eg, combination of therapy via eHealth and in-person care) [38] can provide new treatment possibilities and make it easier for patients to follow up on their own treatment between consultations [39]. An advantage of such blended care is that the health care provider and patient may plan the course of treatment together and that the provider may guide the patient when needed. As such, blended care may offer treatment that can be effective and, at the same time, cost-effective in terms of fewer health care sessions and may lower dropouts because of the individualized follow-up, and the continued access to the eHealth intervention may potentially also help maintain achieved long-term changes [40].

Patients with chronic pain and those providing treatment and care services to these patients have displayed an interest in, and a positive attitude toward, using eHealth self-management interventions for chronic pain [41,42]. Still, blended care in terms of PMP combined with eHealth self-management interventions is yet to be implemented and studied in clinical practice in Norway.

Objectives

The aim of this study is to explore how patients with chronic pain experience the combination of PMP and the use of EPIO, an eHealth self-management intervention for chronic pain.

Methods

Study Design and Participants

This study is part of a larger project aiming to design, develop, pilot-test, and examine the effectiveness of EPIO, a digital self-management intervention for patients living with chronic pain in Norway. The design, development, and feasibility pilot-testing of EPIO have been published elsewhere [41-44], and a large-scale randomized controlled efficacy trial is currently ongoing.

This study adds to the larger project through reporting on qualitative findings from individual interviews conducted to explore patients' experiences using EPIO in combination with PMP. The study was conducted at a physiotherapy institute in Norway. Recruitment processes and study execution was conducted by the first author, who is a psychomotor physiotherapist. To our knowledge, a few, if any, studies have examined delivery of blended care within PMP (eg, digital self-management in combination with PMP), and conducting a small-scale study was, therefore, considered appropriate to test delivery forms and study procedures. With options for the

intervention to last up to 16 weeks (description of PMP and the EPIO intervention program is given later), good opportunities for in-depth descriptions from the participants regarding their experiences of using EPIO in combination with PMP were anticipated. Given the qualitative nature of the study, a predefined number of 6 participants representing patients living with chronic pain and receiving PMP was defined as acceptable (ie, minimum), as even a limited number of participants can allow for richness and depth in data related to participants' experiences (ie, provided data collection methods).

Even with a small number of participants, the study aims for purposively sampling [45] to ensure some variations in gender and pain-related diagnoses. Participants were recruited between May and August 2019 and had to be aged ≥18 years, be able to understand and speak Norwegian, and had to have access to a smartphone or tablet with web access. Exclusion criteria were self-reported untreated severe mental illness, migraine, or cancer-related pain. Potential participants, already enrolled in a PMP treatment program at the physiotherapy institute, were identified based on the inclusion or exclusion criteria and informed about the study by the first author. As the first author was also the psychomotor physiotherapist of the potential participants, recruitment processes were carefully reviewed and discussed by the research team before recruitment to ensure attention to ethical considerations.

The PMP Treatment Program

The PMP treatment program for people with chronic pain aims to facilitate change by raising insight into and awareness about the relationships between experiences and bodily pain [46]. Specifically, the psychomotor physiotherapist in this study worked to help facilitate changes related to bodily tensions, including aiming to strengthen coping mechanisms and reduce tension and pain in muscles, tendons, and joints. This was done by teaching specific movements and exercises and providing massages, all in parallel with seeking to increase patients' knowledge and understanding related to physical reactions and the expression of such.

Part of the idea behind PMP is that people, often without real awareness of doing so, may use their posture, respiration, and muscles to *lock* feelings related to situations of conflict. By *pulling themselves together*, they may end up with tight muscles and subsequent related aches and pains [47]. In the PMP treatment program, the goal was to help raise insight into and awareness about these connections and reduce tension and facilitate change and ultimately to better understand potential *mind-body* connections so the patient could recognize their body as *one functional unit*.

The EPIO Intervention Program

EPIO (inspired by the Greek goddess for the soothing of pain, Epione) is an app-based intervention program designed and developed by a collaborative research team consisting of scientists, health care providers, eHealth experts, content and system developers, and end user representatives [44]. The content is based on well-known, evidence-based aspects from CBT and ACT [28,29] and input from patients, spouses [41], and health care providers [42]. The intention of the EPIO program is to support self-management and coping for patients living with chronic pain.

The EPIO program consists of 9 modules with informational and educational topics, including a variety of self-management–based coping techniques, such as breathing and relaxation exercises, cognitive restructuring exercises, and mindfulness-based information and training. The nine modules are (1) information about pain; (2) balance; (3) thoughts and feelings; (4) stress and coping; (5) what is important to me; (6) health behaviors and lifestyle; (7) communication, relation, and social support; (8) coping during difficult times; and (9) summary and the road ahead. Modules 1 to 5 must be opened in the presented order, whereas from 6 to 8, the user can choose what to open next. To encourage practice, each module must be completed and opened for 3 days before the next module can be opened. While working with the program, users can create their own favorite list by highlighting information and exercises they like, and they can also choose to receive reminders according to their own needs. Users can also choose between reading and listening to the program at any time. To ensure availability, the program can also be used in off-line mode. Screenshot examples from the EPIO intervention program are shown in Figure 1. The EPIO program has so far been tested in a feasibility pilot study with 50 participants, where the results showed EPIO to be perceived as useful and easy to use, with easily understandable content and excellent system usability [43].

When enrolled in the study, participants were introduced to the self-management program and received help downloading the EPIO app from App Store (Apple Inc) or Google Play Store, with instructions on how to get started. They were informed that they could use EPIO as much as they wanted in the study period of a maximum of 16 weeks. As such, they had access to a program containing some well-known but also many new exercises and aspects of information that they could use between PMP treatment sessions, potentially bringing new aspects into the treatment. When coming in for PMP treatment, the first part (ie, 5-10 minutes) of the face-to-face session with the psychomotor physiotherapist was spent talking about patients' experiences of using EPIO since the last PMP session. They were given the opportunity to ask questions or introduce topics they had been involved in through EPIO or tell if there was something special that occupied them, thereby delivering EPIO in a blended care type of model.

Figure 1. Example screenshots from the EPIO intervention program.
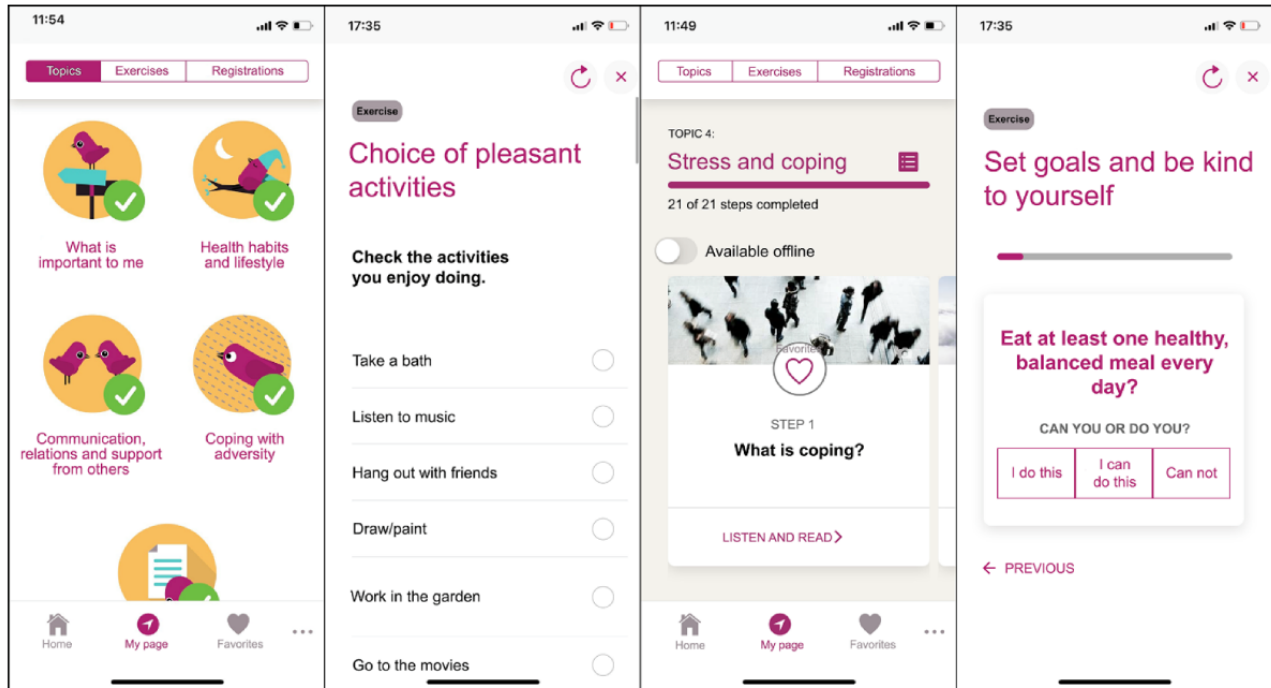
Data Collection Procedure

Information about age, diagnosis, and years of living with chronic pain was collected using a study-specific demographic form at study enrollment. Semistructured individual interviews were conducted following the completion of all modules in the EPIO program. A study-specific interview guide was developed based on literature examining and describing self-management of chronic pain, digital self-management interventions, and blended care health care delivery [4,24,27,30,35,36,39]. The interview guide contained questions about EPIO use and usefulness, pain impact, coping strategies, and potential experiences of change based on the use of EPIO (the interview guide is given in Multimedia Appendix 1). The questions were developed with the intention of allowing the participants to give a free description of their experiences. The interviews were conducted by the first author at her office from September to December 2019.

Each interview started with an open-ended question about the participant's experience with using EPIO. On the basis of the answers, either new topics were addressed for follow-up or the next topic from the interview guide was introduced. The interviewer did not follow a specific order for the topics to be addressed but rather focused on having a fruitful conversation where topics were introduced and they appeared to fit naturally into the conversation, focusing on topics and hints addressed by the participant to capture issues raised. Furthermore, the interviewer asked clarifying questions to ensure that the essence of what the respondents described was clearly understood. The interviews lasted between 35 and 65 minutes, depending on the richness of the conversation. The interviews were audio recorded and transcribed verbatim, resulting in a total of 60 densely written text pages for analysis.

Data Analysis

The transcripts were analyzed using systematic text condensation in 4 steps [48]. In step 1, the first author read through the transcripts several times while taking notes to get an overview of content and meaning. To strengthen the data analysis process, coauthors DN and CV also read through some of the transcripts to become familiarized with the material and prepare for further team analysis and discussions. In step 2, the first author identified meaningful entities (encoding) in collaboration with coauthors DN and CV and coded the identified material in the transcripts into 4 overarching themes using NVivo software (QSR International). In this process, the authors first identified preliminary themes, which were then discussed and negotiated into a final set of themes. In step 3, the 3 authors performing the analysis created subthemes and abstracted the content of the individual meaningful entities into each subtheme in the condensates. In step 4, the meaning of the condensates was summarized. The themes and subthemes were reviewed and rearranged into a final structure by AGE, DN, and CV. The authors met regularly throughout the analysis process, considering different ways of interpreting the results, and reached consensus through discussions.

Finally, the themes were approved before the authors selected representative quotes to support the analysis. The full transcripts were not translated from Norwegian to English. However, the third author (LSN), who is bilingual (ie, Norwegian or English), was well familiar with the transcripts and ensured that the translation of quotes and other aspects needed for interpretation preserved the original meaning of the transcripts and quotes.

Ethical Aspects

The study was planned and performed in compliance with the principles outlined in the Declaration of Helsinki [49]. The study was approved by the Regional Committee for Medical and Health Research Ethics (REK 2018/8911) and the Oslo

University Hospital Department for Data Protection and Information Security (ie, institutional review board equivalent; PVO 2017/6697). Written informed consent was obtained from all participants.

To ensure anonymity and privacy of the participants, personal data (eg, contact information and date of birth) and signed consent forms were stored on a secure server at Oslo University Hospital, separately from the collected data from the interviews. Each participant was assigned a study identity number, and the codebook connecting this identity number to the participant was stored separately on the secure server. During the audio-to-text transcription process, all names were removed, the data were deidentified, and the analysis was conducted using the deidentified transcribed text only.

Results

Participant Information

A total of 6 patients were approached and agreed to participate in the study. Of the 6 participants, 1 (16.7%) did not complete the program because of personal life events and was therefore not interviewed after the intervention and 5 (83.3%; 4 women and 1 man) completed the program and participated in the interviews. These participants spent 10 to 15 weeks from the time they received access to the EPIO program until all the 9 modules were completed, and the interviews were conducted. During the same period, the participants had 7 to 13 PMP face-to-face sessions. The patients who completed the program were aged between 30 and 70 years, with an average age of 51 (median 52) years. All patients had lived with pain for more than 3 months, most of them for many years (ie, mean 12.4, range 2-40 years). Diagnoses included pain after trauma or postaccident traumas, neck pain, and diffuse pain in the musculoskeletal system. Other diagnoses included posttraumatic stress disorder, generalized anxiety disorder, and attention-deficit/hyperactivity disorder. All participants reported struggling with the impact of pain on various areas of their daily lives. The familiarity with the use of mobile apps varied from moderate to very experienced, with 4 of the patients completing the program reported having previously used other health-related mobile apps.

Overview

Transcripts from the interviews were classified into four main themes: (1) awareness, (2) relationship to oneself, (3) relationship to pain, and (4) PMP and EPIO combined. The main themes and subthemes are illustrated in Figure 2.

Figure 2. Overview of the main themes and subthemes. PMP: psychomotor physiotherapy.
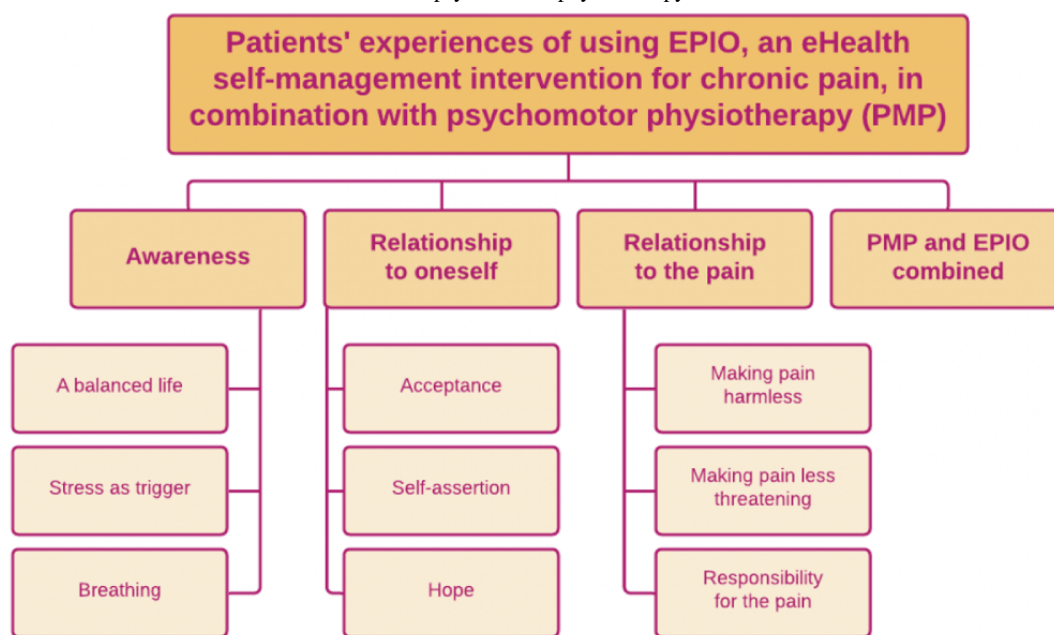
Awareness

Overview

All participants reported increased awareness about how they lived their lives and how they have become more attentive toward their body after using EPIO combined with PMP. The changes described related to awareness were sorted into three subthemes: a balanced life, stress as a trigger, and breathing. The participants described changes in how they noticed and reacted to signals from their body and described new thoughts about balance in life as a key to being able to live better with pain. All participants also reported increased awareness of the causes of stress and ways to better handle stress.

A Balanced Life

All participants described how EPIO had helped them to consider their own balance of rest and activity, and most of them stated that they, because of EPIO, had become aware of how their activity level was often higher than what they had thought. Some of the participants described becoming aware of a gap between how they wanted, and tried, to live their lives and the extent to which they were actually able to live their lives this way. Explaining how EPIO had helped, one participant stated the following:

I have been 'on' for as long as it is possible to be "on". Until I'm exhausted, and then I am "off". So, to keep it more even during the days, it is easier to remind myself now that I have this more at the forefront of my mind. [Participant 5]

A number of the participants also stated that they had discovered how important it was to take short breaks and be conscious of their level of activity, rest, and sleep. Some reported feeling calmer after using EPIO, and some reported that they used EPIO as a reminder to take breaks and perform EPIO exercises throughout the day. A participant said this about how EPIO gave concrete suggestions to balance activity and rest:

But this was very specific, you know, can you do half the dishes and then rest? I never even thought of it that way. [Participant 4]

Stress as a Trigger

All participants stated that the use of EPIO had made the relationship between stress and pain clearer to them and that they realized they had to do something about the stress and the way they dealt with stress in their lives. A participant said the following:

It was quite early [after being introduced to EPIO], that I had to realize the importance of stress-reducing approaches. [Participant 3]

Participants did, however, present different stories about the causes and nature of their stress reactions. Some described stress at work or at home, whereas others described having been exposed to life trauma causing bodily and psychological stress. Several participants stated that EPIO helped them gain an increased awareness of the origin of their stress responses and had also provided tools for better self-management. A participant stated the following:

To realize that my breathing is connected to stress and traumas and improvement has been surprising. When I get flashbacks [...], this is what happens. My body goes into alert mode. [...] Sometimes it takes time to calm it down, sometimes it goes more quickly. But I HAVE a tool, to get my body out of alert mode. And it is the breathing techniques from the app. [Participant 4]

Participants also described often finding it challenging to make the changes needed to lower their stress levels, stating that taking frequent breaks and doing breathing exercises or actively using approach coping strategies can be easier said than done during demanding or stressful times. A number of them stated that it may take time and effort to start using such strategies frequently and that such suggested changes and approaches may sometimes seem impossible when an actual stressful situation occurs. However, all participants described the face-to-face contact with, and follow-up by, the psychomotor physiotherapist as an advantage when aiming to make the necessary changes. One participant reported realizing their home situation was their main source of stress, with years of *serving* other family members making the participant feel as if having almost *disappeared* as an individual. For the road ahead, the participant shared the following:

Having insight, or taking time for yourself, it's kind of...It hasn't really been there before, and it opens a few doors that can create a bit of a mess around you. Not just for yourself, but for the people around you too [...] Getting insight related to the fact that I do things for others all the time, for everyone else...It can make me irritated, and maybe a little bit angry [...] I haven't been angry for years. Maybe this is a good thing, there is after all a reason for all the tensions. [Participant 3]

The quote illustrates how challenging, and even stressful, it can be for a person to make the necessary suggested changes and how it can also impact people around them and their relationship with them.

Breathing

All participants described experiencing becoming more aware of, and attentive to, their bodies through the use of EPIO. The focus on breathing and diaphragmatic breathing techniques in EPIO was reported to be among the most important parts of the program, with the breathing focus reportedly bringing on reflections about, and experiences of, how deep breathing could help make the mind as well as the body more relaxed. Some of the participants had very specific experiences of how breathing could impact their pain, and one stated the following:

It was shocking to realize how breathing specifically helps with tolerating and coping with pain. I'm thinking, it hurts, I will try using the EPIO app now. Then I do one of the exercises, and I notice, it hurts to breathe, but I can see that...When I do that exercise, and manage to breathe properly, the pain eases...It eases and becomes more tolerable. [Participant 4]

Most participants described finding value in receiving information about the importance of deep breathing, stating that this created a new awareness about their own breathing and tension patterns. They also emphasized the importance of being guided through breathing and relaxation exercises by the EPIO voice, to feel the difference guidance could make, and the importance of being encouraged to repeat the exercises on their own over a period. With this new awareness related to the role of breathing, participants described the EPIO exercises as helpful in breaking old patterns. A participant stated the following:

My body can find relaxation without me forcing it to. I don't have to force my shoulders to relax in order to breathe. I can just think about it, and balance occurs. It has strengthened my belief, or convinced me, that things are connected. There is no doubt. Our mind controls a lot of what we do, whether we are aware of it or not. The fact that you can actually lead yourself out of what has just occurred - without even really being aware. [Participant 1]

Relationship to Oneself

The participants reported that EPIO had contributed to changes in how they related to themselves, with some of them even referring to EPIO as a program for self-management of life in general, sometimes even forgetting that the program was about pain and pain management. The changes described as

experienced by the participants were sorted into three subthemes: acceptance, self-assertion, and hope.

Acceptance

Some participants described that EPIO had helped them gain greater self-acceptance, stating that they now felt more confident, with a stronger belief in themselves, realizing that they were not *a bad person*, even if the pain kept them from doing and managing things the way they had previously been able to. A participant stated the following:

> *Some of the things I have learned using EPIO have made it easier for me to accept life the way it is. And to understand it, and not beat myself up about it.* [Participant 1]

In addition, some of the participants found it important to consider the qualities and skills they still had, what they wished for in life, and what they appreciated in the people they were close to. A participant said the following:

> *And, I think, it's not like I have to reinvent myself. [...] No, I AM there! [...] So the value, my personality, is not gone. The skills are there, but I cannot do as much as I did before.* [Participant 4]

Self-assertion

Several of the participants described feeling empowered by some of the content in the EPIO program, stating that EPIO encouraged them to assert themselves. Some of the participants described finding time to do what was important and right for themselves as challenging and unfamiliar, stating that prioritizing themselves and asking for help appeared unfamiliar and had not had precedence in their lives. A participant said the following:

> *I am the person helping others. No one thinks I should need help. So I really have myself to thank for that. People do not come and ask.* [Participant 5]

However, some of the participants described appreciating being reminded of the importance of taking time for themselves, prioritizing their own health, wishes, and needs. A participant even related his or her pain to the lack of prioritizing themselves and stated the following:

> *The pain in my muscles, it can of course come from the fact that I live to help others all the time, and forget to help myself in a way. And that's in this program. Help with self-management.* [Participant 3]

Hope

Some of the participants also reported having obtained new hope in life from the use of EPIO in combination with PMP. They described having gained hope of being able to manage life in a new and different way, including hope of being able to live a more balanced life, to know and actively use coping strategies in day-to-day life, and through that being able to live life, even with pain—in a more fulfilling and *normal* way. They described being able to have a tool such as EPIO, accessible anywhere and anytime, as providing them hope for better quality of life and a better future ahead. A participant described the link between EPIO and new hope this way:

> *This program has given me hope that I can continue to work on this. I will have a life with quality, I will have a life I can manage. I will live a life I don't have to be ashamed of. I will be a complete person, but I will do things a little differently, and at a different pace.* [Participant 4]

Relationship to the Pain

The participants described a number of ways that EPIO had contributed to changes in how they related to their pain. Descriptions of such changes related to experiences that EPIO had contributed to making the pain less threatening, encouraged and stimulated the participants to use distractions from the pain, and changed their view of their own responsibility in terms of managing the pain.

Making Pain Less Threatening

Most participants reported that they, after starting to use the EPIO program, had experienced less pain-related anxiety and described EPIO as helpful in making the pain seem less threatening in their day-to-day life. To some extent, it had been important to gain new insights into physiological explanations and pain coping techniques. They stated that the information and knowledge gained through EPIO made it easier to relate to potential causes of pain and understand the relationships between pain, thoughts, and feelings and that the information gained through EPIO made it easier to believe that the pain might not necessarily be dangerous. Learning new coping strategies also appeared to have contributed to making the pain seem less threatening, as expressed by a participant:

> *I can cope better with my pain if I know I have techniques I can use during the day to make it a little less tense and painful. Yes, the pain is still there, but...It does not get as much attention.* [Participant 2]

Distraction From the Pain

The EPIO program contains a number of coping strategies and exercises related to distraction from the pain, and all participants in the study described learning and using these techniques, aiming to distract and remove the focus from pain, as helpful. Some of the participants even explained that because of deliberately using these distraction techniques, they now had periods where the pain appeared to absorb less of their day-to-day attention. Others stated that the pain was still present but that it might be less dominating. One of the participants described the increased ability to distract themselves from the pain as follows:

> *Just to be able to move away from focusing on the pain. That things in my body are not as they should, to just focus on something else for a while. A break. [...] To just notice what happens around you. Not just inside you. Focus on that.* [Participant 2]

Responsibility for the Pain

After using EPIO, several participants stated that discovering strategies to influence and cope with their pain had made them realize that they have a greater responsibility of managing their own pain than what they had previously thought. They all stated

that they actively had to handle the pain and described how they could see more clearly how their pain related to the way they lived their lives, including impact on stress factors, relationships, breathing habits, and bodily tensions. However, most participants stated that making the necessary changes in life would be energy consuming and that establishing lasting changes would require motivation and energy. One participant described making significant changes as "impossible" because of the situation at home. Another participant described the significant work required to obtain the effect as follows:

> Pain is something I have to relate to in one way or another. How I think is very important, and I have to think that I can handle things in a different way. Quick fix, that is what you want [...] And this is the opposite, in a way. You have to really delve into it, absorb it, and let it have consequences for your body. [Participant 1]

However, the same participant added a contrary reflection, noting the possibility that EPIO could potentially produce guilt and feelings of resignation should the user do everything to integrate what they had learned while using the app, without experiencing improvements in pain levels:

> I have felt it myself and, well now I have felt better, and it may be because I have done all the right things. And if you're then suddenly in pain, then it may feel like it's your own fault. [Participant 1]

PMP and EPIO Combined

All participants in this study reported experiencing EPIO and PMP as complementary and stated that they found it easier to perform exercises and focus on pain management strategies while using EPIO in combination with PMP treatment. All participants described EPIO as a reinforcing tool for their pain self-management, bringing new aspects and themes to their PMP treatment, and reported finding it easier to work on their own, in between PMP treatment sessions, with EPIO at hand. They described EPIO as an enforcer, making it easier to integrate the necessary pain self-management into daily life. One of the participants said the following:

> I think this [EPIO] is a very useful tool to use in parallel [with PMP], and the patient can work in between. I think it could help many to integrate it into day-to-day life. [Participant 2]

Several participants also pointed to the combination of self-management through EPIO and face-to-face or physical touch provided by PMP as an important part of the process of raising mental and physical awareness. All participants described the massage techniques learned from PMP as having helped them become more aware of tension in their muscles, making it easier to achieve relaxation. One of the participants, having experienced significant traumas in life, also stated that the combination of PMP and EPIO helped involve the body as well as the mind in the process of treatment and healing.

During the interviews, participants frequently mentioned how making changes can be difficult and sometimes even frightening. As such, having a combination of in-person or EPIO treatment available was seen as a strength, enabling learning at the participants' own pace but with the safety of in-person discussions and guidance when needed. A participant stated the following:

> I think you need a combination. And some time. Because this is something you can choose not to do. With everything else going on in life. It's easier to just sit down and read the paper. And if you are in great pain, you don't have that much to go on. [Participant 5]

As EPIO was described as comprehensive, all participants stated that it was helpful to work together with the psychomotor physiotherapist to work through potential issues that occurred and sort out what was the most important aspect to focus on between each consultation. Several patients also stated that EPIO had started a process in which they had to actively keep working. They all said that they thought EPIO alone would have brought new experiences also without PMP but that the combination strengthened the impact.

Discussion

Principal Findings

This study explored patients' experiences of using the app-based intervention EPIO in combination with PMP for self-management of chronic pain. The study participants described experiencing EPIO as bringing well-known as well as new themes into the treatment and as supporting self-management between face-to-face sessions. Following the combined EPIO and PMP program, participants described increased awareness of a number of issues, including the importance of balancing rest and activity, the relationship between stress and pain, and the importance of being aware of signals from their bodies. They also reported that EPIO had contributed to changes in their relation to themselves, including increased self-acceptance and self-assertion, willingness to prioritize themselves, and increased hope for a better life. Participants' relationship with the pain itself was also reported to have changed when using EPIO in combination with PMP, including experiencing pain as less threatening, benefiting from the use of distractions, and taking more responsibility for self-management of the pain.

Comparison With Previous Work

Findings show that the participants experienced living and coping with pain as complex—involving challenges related to many aspects of life, including stress. Some researchers have suggested a link between stress and hypersensitivity to pain [50,51], and in accordance with existing research [52], participants in this study described how pain was triggered and reinforced by psychosocial stressors. The introduction of EPIO into PMP treatment led to positive changes in how the participants perceived their pain and approached their daily lives. In line with studies showing the interrelation between stress, pain, and coping [51,52], participants described becoming more aware of causes of stress in their daily lives, and how stress affected them and triggered their pain. Through the blended treatment of EPIO and PMP, most participants reported

having found new strategies to deal with life stress and better manage their pain.

The participants described psychosocial stressors related to social relationship challenges, feelings of guilt, difficulties meeting the many demands of daily life, and a general feeling of not being enough and constantly falling short of demands and expectations. Research has suggested that people living with chronic pain may benefit from treatment approaches aimed at reducing feelings of guilt, finding ways to cope with the many demands, and increasing levels of confidence [53]. Participants in this study reported becoming more self-accepting, assertive, and confident and feeling less shameful and vulnerable, which may therefore be of importance in this setting.

The interrelationship between fear, anxiety, and pain is commonly reported among patients with chronic pain, and scholars emphasize the importance of addressing these aspects in treatment [54,55]. In this study, participants reported having become less fearful and anxious about their pain following the blended EPIO and PMP treatment, attributing the changes to the general information about pain and tools for better coping provided by EPIO.

Following the blended EPIO and PMP treatment, participants in this study described having become more aware of breathing in general and the importance of deep breathing in particular. The importance of breathing is a core part of PMP treatment and had already been introduced to participants before the study. Still, some exercises in the EPIO program reportedly provided new experiences regarding the relationship between deep breathing and pain management, apparently raising new awareness of these relationships, and some of the participants reported experiencing reduced levels of stress and pain because of practicing breathing technique exercises. As deep, diaphragmatic breathing can impact breathing frequency (ie, slowing), these findings can also support the notion that associations exist between breathing frequency and pain [56].

Participants described a new awareness of the relationship between rest and activity. Some reported realizing that their activity level had reached beyond their capacity, and that this had been the case for quite some time. These participants described it as difficult to take breaks during the day and ask for assistance. Their observation of the importance of a better balance between rest and activity corresponds well with research reporting that imbalance can lead to reduced pain tolerance and decreased quality of life [57,58].

Self-management tools such as EPIO generally aim to help patients take an active role in managing their pain and day-to-day life. Self-efficacy is often seen as a determinant for keeping up efforts and not giving in when faced with obstructions [59], and findings from this study may indicate an increased sense of self-efficacy among participants posttreatment, in terms of a strengthened belief in being able to live a life less impacted by pain. A strengthened sense of self-efficacy may also have helped them make the necessary changes in life to achieve better coping.

Self-management can be demanding, entailing taking responsibility to *manage* one's own life, maintaining good

habits, and making changes when needed. To make the necessary changes requires energy and motivation, of which the patient may already be drained or depleted because of the many demands involved in living with chronic pain [1,12,60]. Although participants reported gaining new awareness and knowledge related to pain and self-management of pain from the EPIO program, they all underlined the demands and costs of being the agent of self-management and change in this way. EPIO was not considered *a quick fix* either, as the program requires users to be active and engaged throughout. Self-management programs such as EPIO may help users discover what they can or need to do to better live *with* pain. However, if unable to undertake the necessary steps toward better self-management, knowing that life could be better *if only* changes are made may possibly also produce feelings of guilt and shame, making life even worse. This implies responsibility and developers and/or providers of such self-management programs should be aware of the potential ethical implications of creating and recommending such independent self-management programs, regardless of potential benefits and intent. Participants in this study described appreciating the combination of EPIO and PMP in their treatment, and it could be that self-management programs such as EPIO could best be recommended in combination with existing care and in-person follow-up, as also supported by the participants in this study. However, several participants reported having gained increased awareness of their own responsibility for self-management in this study, which they were also willing to take. It is possible that, despite the demanding aspects of self-management, the CBT and ACT basis of EPIO may have helped the participants achieve a changed relation to their pain, pointing them toward a greater acceptance of living with pain and an enhanced knowledge of what could be done.

The use of blended care, here referred to as combining in-person and eHealth programs, may have a positive impact as patients may benefit from the best of 2 or more treatment approaches [38,39]. Having EPIO as a tool to work with between PMP sessions appeared to help the participants in this study, and this type of blended care may be suitable for stress- and pain-related self-management and coping, as indicated in this study.

With more than one treatment option available (eg, PMP and EPIO), patients can have the option to decide for themselves how to meet their treatment needs, for example, from their psychomotor physiotherapist or from a program such as EPIO or both. In line with previous research, the results from this study indicate that many prefer to go through the eHealth program at their chosen pace [61], without a health care provider or coach and guide trying to impact their progress. Furthermore, compared with in-person interventions, eHealth interventions can be experienced as a more neutral and nonjudgmental form of delivery, eliminating the fear of being judged or misunderstood by a provider [61]. This corresponds with participants' description of finding it easier to go through the contents of EPIO on their own, without necessarily discussing everything with the psychomotor physiotherapist. At the same time, some of the participants brought relevant topics from their EPIO use into psychomotor therapy; topics they may not have discovered had the information been less comprehensive.

Another important reason for combining EPIO and PMP is the notion that the patient–provider relationship can affect patient outcomes. A strong therapeutic alliance between the patient and health care provider can potentially predict positive treatment outcomes [62,63]. In addition, speaking to another person can be helpful. Simultaneously, if the therapeutic alliance is weak, the treatment process may become more difficult. In such cases, an option such as EPIO can be particularly useful. eHealth can likely not replace the relational processes involved in in-person treatment. However, eHealth interventions can offer support when in-person treatment is not available, and at the very least, eHealth can be offered as a supplement to in-person treatment and care, as in this study. Through EPIO, the participants in this study received a CBT and ACT approach beyond PMP regular competence, describing the information and content in EPIO as meaningful and complementary to PMP treatment.

Strengths and Limitations

This study has a number of limitations. First, the study was conducted at a single physiotherapy institute and included a limited number of participants. However, although the sample was small, the data material represented richness and depth about how the participants experienced using EPIO in combination with PMP, and statements from one participant could be as important as statements of most participants. Given the small predefined number of participants in the study, all participants were identified and selected by the psychomotor physiotherapist to ensure that the participants could use EPIO in the study period and be able to attend a postintervention interview. This could have introduced a selection bias and as such be considered a limitation, but it also ensured professional soundness (ie, able to attend the study). Although conducted with a limited number of participants, the interviews conducted in this study provided insight into a variety of patient experiences and allowed sufficient depth in the analyses, which can be considered a strength.

Second, 80% (4/5) women and 20% (1/5) men participated in this study. The results of the study might have been different if more men had participated. For example, men might bring different experiences of using EPIO in combination with PMP. However, the gender distribution in the study represents the general gender distribution among patients receiving PMP, and additional men were not available for inclusion at the time the study was conducted. This is also in line with the first author's experience from clinical practice, with approximately 1 man per 12 to 15 women enrolled to receive PMP. In addition, given the range related to age, years living with pain and other diagnoses, and most women living with chronic pain, the study can be considered representative of patients living with chronic pain. In addition, with the rich data material collected during and following study participation, even a small number of participants allows for in-depth insight related to participants' experiences using EPIO in combination with PMP.

As 80% (4/5) participants in the study had previously used health-related mobile apps, a certain bias related to previous eHealth experience could be present in the current findings. Such experiences could impact participants' opinions and views, and perceived benefits and findings might have looked different if the sample included a more balanced sample of experienced or inexperienced users. Experience in using digital tools could, however, impact findings positively (eg, have enjoyed using mobile apps) and negatively (eg, did not like using mobile apps), and the outcome of having a more balanced sample may therefore be challenging to predict.

Participants were also identified and selected by the first author and were already enrolled in the first author's PMP treatment program. This may have implied a selection bias but may also have ensured professional soundness (ie, patients who were able to complete the EPIO program). To ensure that patients did not feel coerced to participate in the study, they received oral and written information stating that study participation was voluntary; nonparticipation would have no consequences for their ongoing PMP treatment; and they, should they decide to participate, could withdraw from the study at any time without having to give a reason. Patients received information about the study in one PMP session and could then wait until a later time point (ie, time to reflect) to decide whether to participate and give their written consent. To limit the risk that participants provided more positive than negative feedback because of their relationship with the psychomotor physiotherapist, the participants were informed that study results would not affect the psychomotor physiotherapist, the ownership of EPIO was elsewhere, and the psychomotor physiotherapist did not receive any benefits or have any negative consequences, regardless of the study findings. Before the study interviews, participants were also informed that to truly evaluate the program, their honest feedback on EPIO, and about using EPIO in combination with PMP, was needed.

Throughout the study process, the authors adopted a reflexive approach to the dual role of the researcher or psychomotor physiotherapist. This was a point of discussion in all research meetings throughout the process and was also addressed by closely involving two of the coauthors in the data analysis process.

Third, the first author's professional background as a psychomotor physiotherapist could have influenced the process (ie, from interviews to manuscript write-up) through preconceptions drawing attention to certain aspects of information. The involvement of 2 coresearchers during analysis was, however, actively used to counter such threats to the trustworthiness of the findings [64].

Finally, the results showed that living with pain is overly complex, and a small qualitative study such as this cannot provide information about what contributed the most to participants' experienced benefits of using EPIO or the combination of EPIO and PMP. As such, factors unrelated to the EPIO tool may have contributed, including regular follow-up by a health care provider and the ability to talk about EPIO at every consultation. The close follow-up may also have contributed to the participants making a greater effort with EPIO than those who did not have the blended care type follow-up. The qualitative in-depth insight provided, even with a small sample, is a major strength of this study.

XSL·FO

RenderX

Conclusions

This study offers insight into chronic pain patients' experiences with the use of EPIO, an app-based intervention for the self-management of chronic pain, in combination with PMP. Findings indicate that EPIO, used in this blended care context, was experienced as beneficial in raising awareness and finding a better balance between rest and activity; feeling more valuable as a person; and increasing self-acceptance, self-assertion, and hope and in relation to the participants' pain—seeing pain as less harmful and engaging in more active stress management and coping strategies. In conclusion, this study indicates that a blended care approach combining eHealth self-management interventions such as EPIO with PMP may be of value to patients living with chronic pain.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File , 12 KB - formative_v6i3e34458_app1.docx]

References

1. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain 2006;10(4):287-333. [doi: 10.1016/j.ejpain.2005.06.009] [Medline: 16095934]

2. Landmark T, Romundstad P, Dale O, Borchgrevink PC, Kaasa S. Estimating the prevalence of chronic pain: validation of recall against longitudinal reporting (the HUNT pain study). Pain 2012;153(7):1368-1373. [doi: 10.1016/j.pain.2012.02.004] [Medline: 22575226]

3. Løyland B, Miaskowski C, Wahl AK, Rustøen T. Prevalence and characteristics of chronic pain among long-term social assistance recipients compared to the general population in Norway. Clin J Pain 2010;26(7):624-630. [doi: 10.1097/AJP.0b013e3181e0de43] [Medline: 20639732]

4. Loeser JD, Bonica JJ. Bonica's management of pain. Philadelphia: Lippincott Williams & Wilkins; 2001.

5. Eriksen J, Jensen MK, Sjøgren P, Ekholm O, Rasmussen NK. Epidemiology of chronic non-malignant pain in Denmark. Pain 2003;106(3):221-228. [doi: 10.1016/S0304-3959(03)00225-2] [Medline: 14659505]

6. Hasselström J, Liu-Palmgren J, Rasjö-Wrååk G. Prevalence of pain in general practice. Eur J Pain 2002;6(5):375-385. [doi: 10.1016/s1090-3801(02)00025-3] [Medline: 12160512]

7. Dueñas M, Ojeda B, Salazar A, Mico JA, Failde I. A review of chronic pain impact on patients, their social environment and the health care system. J Pain Res 2016;9:457-467 [FREE Full text] [doi: 10.2147/JPR.S105892] [Medline: 27418853]

8. Landmark T, Romundstad P, Dale O, Borchgrevink PC, Vatten L, Kaasa S. Chronic pain: one year prevalence and associated characteristics (the HUNT pain study). Scand J Pain 2013;4(4):182-187. [doi: 10.1016/j.sjpain.2013.07.022] [Medline: 29913652]

9. Langley PC. The prevalence, correlates and treatment of pain in the European Union. Curr Med Res Opin 2011;27(2):463-480. [doi: 10.1185/03007995.2010.542136] [Medline: 21194390]

10. Rice AS, Smith BH, Blyth FM. Pain and the global burden of disease. Pain 2016;157(4):791-796. [doi: 10.1097/j.pain.0000000000000454] [Medline: 26670465]

11. Macfarlane GJ, Barnish MS, Jones GT. Persons with chronic widespread pain experience excess mortality: longitudinal results from UK Biobank and meta-analysis. Ann Rheum Dis 2017;76(11):1815-1822. [doi: 10.1136/annrheumdis-2017-211476] [Medline: 28733474]

12. Solberg Nes L, Roach AR, Segerstrom SC. Executive functions, self-regulation, and chronic pain: a review. Ann Behav Med 2009;37(2):173-183. [doi: 10.1007/s12160-009-9096-5] [Medline: 19357933]

13. Breivik H, Eisenberg E, O'Brien T, OPENMinds. The individual and societal burden of chronic pain in Europe: the case for strategic prioritisation and action to improve knowledge and availability of appropriate care. BMC Public Health 2013;13:1229 [FREE Full text] [doi: 10.1186/1471-2458-13-1229] [Medline: 24365383]

14. Kelly GA, Blake C, Power CK, O'keeffe D, Fullen BM. The association between chronic low back pain and sleep: a systematic review. Clin J Pain 2011;27(2):169-181. [doi: 10.1097/AJP.0b013e3181f3bdd5] [Medline: 20842008]

15. IsHak WW, Wen RY, Naghdechi L, Vanle B, Dang J, Knosp M, et al. Pain and depression: a systematic review. Harv Rev Psychiatry 2018;26(6):352-363. [doi: 10.1097/HRP.0000000000000198] [Medline: 30407234]

16. Zhao SS, Miller N, Harrison N, Duffield SJ, Dey M, Goodson NJ. Systematic review of mental health comorbidities in psoriatic arthritis. Clin Rheumatol 2020;39(1):217-225. [doi: 10.1007/s10067-019-04734-8] [Medline: 31486931]

17. Grabovac I, Dorner TE. Association between low back pain and various everyday performances : activities of daily living, ability to work and sexual function. Wien Klin Wochenschr 2019;131(21-22):541-549 [FREE Full text] [doi: 10.1007/s00508-019-01542-7] [Medline: 31493101]

18. Malmberg-Ceder K, Haanpää M, Korhonen PE, Kautiainen H, Soinila S. Relationship of musculoskeletal pain and well-being at work - does pain matter? Scand J Pain 2017;15:38-43. [doi: 10.1016/j.sjpain.2016.11.018] [Medline: 28850343]

19. Nordstoga AL, Vasseljen O, Meisingset I, Nilsen TI, Unsgaard-Tøndel M. Improvement in work ability, psychological distress and pain sites in relation to low back pain prognosis: a longitudinal observational study in primary care. Spine (Phila Pa 1976) 2019;44(7):E423-E429. [doi: 10.1097/BRS.0000000000002860] [Medline: 30234815]

20. Mills S, Torrance N, Smith BH. Identification and management of chronic pain in primary care: a review. Curr Psychiatry Rep 2016;18(2):22 [FREE Full text] [doi: 10.1007/s11920-015-0659-9] [Medline: 26820898]

21. Turk DC, Wilson HD, Cahana A. Treatment of chronic non-cancer pain. Lancet 2011;377(9784):2226-2235. [doi: 10.1016/S0140-6736(11)60402-9] [Medline: 21704872]

22. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness: the chronic care model, part 2. JAMA 2002;288(15):1909-1914. [doi: 10.1001/jama.288.15.1909] [Medline: 12377092]

23. Nicholas MK, Blyth FM. Are self-management strategies effective in chronic pain treatment? Pain Manag 2016;6(1):75-88. [doi: 10.2217/pmt.15.57] [Medline: 26678703]

24. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]

25. Dansie EJ, Turk DC. Assessment of patients with chronic pain. Br J Anaesth 2013;111(1):19-25 [FREE Full text] [doi: 10.1093/bja/aet124] [Medline: 23794641]

26. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]

27. Dragesund T, Øien AM. Demanding treatment processes in Norwegian psychomotor physiotherapy: from the physiotherapists' perspectives. Physiother Theory Pract 2019;35(9):833-842. [doi: 10.1080/09593985.2018.1463327] [Medline: 29683774]

28. Beck AT. Cognitive therapy and the emotional disorders. New York: Penguin Books; 1976.

29. Hayes SC, Strosahl KD, Wilson KG. Acceptance and commitment therapy: an experiential approach to behavior change. New York: The Guilford Press; 1999.

30. Morley S, Eccleston C, Williams A. Systematic review and meta-analysis of randomized controlled trials of cognitive behaviour therapy and behaviour therapy for chronic pain in adults, excluding headache. Pain 1999;80(1-2):1-13. [doi: 10.1016/s0304-3959(98)00255-3] [Medline: 10204712]

31. de C Williams AC, Fisher E, Hearn L, Eccleston C. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database Syst Rev 2020;8(8):CD007407. [doi: 10.1002/14651858.CD007407.pub4] [Medline: 32794606]

32. Kashdan TB, Rottenberg J. Psychological flexibility as a fundamental aspect of health. Clin Psychol Rev 2010;30(7):865-878 [FREE Full text] [doi: 10.1016/j.cpr.2010.03.001] [Medline: 21151705]

33. Hann KE, McCracken LM. A systematic review of randomized controlled trials of Acceptance and Commitment Therapy for adults with chronic pain: outcome domains, design quality, and efficacy. J Contextual Behav Sci 2014;3(4):217-227. [doi: 10.1016/j.jcbs.2014.10.001]

34. Veehof MM, Trompetter HR, Bohlmeijer ET, Schreurs KM. Acceptance- and mindfulness-based interventions for the treatment of chronic pain: a meta-analytic review. Cogn Behav Ther 2016;45(1):5-31. [doi: 10.1080/16506073.2015.1098724] [Medline: 26818413]

35. Heapy AA, Higgins DM, Cervone D, Wandner L, Fenton BT, Kerns RD. A systematic review of technology-assisted self-management interventions for chronic pain: looking across treatment modalities. Clin J Pain 2015;31(6):470-492. [doi: 10.1097/AJP.0000000000000185] [Medline: 25411862]

36. Thurnheer SE, Gravestock I, Pichierri G, Steurer J, Burgstaller JM. Benefits of mobile apps in pain management: systematic review. JMIR Mhealth Uhealth 2018;6(10):e11231 [FREE Full text] [doi: 10.2196/11231] [Medline: 30348633]

37. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the mobile app rating scale. JMIR Mhealth Uhealth 2018;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]

38. Massoudi B, Blanker MH, van Valen E, Wouters H, Bockting CL, Burger H. Blended care vs. usual care in the treatment of depressive symptoms and disorders in general practice [BLENDING]: study protocol of a non-inferiority randomized trial. BMC Psychiatry 2017;17(1):218 [FREE Full text] [doi: 10.1186/s12888-017-1376-1] [Medline: 28610561]

39. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JE. Mixing online and face-to-face therapy: how to benefit from blended care in mental health care. JMIR Ment Health 2016;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]

40. Erbe D, Eichert HC, Riper H, Ebert DD. Blending face-to-face and Internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]

41. Ledel Solem IK, Varsi C, Eide H, Kristjansdottir OB, Mirkovic J, Børøsund E, et al. Patients' needs and requirements for eHealth pain management interventions: qualitative study. J Med Internet Res 2019;21(4):e13205 [FREE Full text] [doi: 10.2196/13205] [Medline: 30877780]

42. Varsi C, Ledel Solem IK, Eide H, Børøsund E, Kristjansdottir OB, Heldal K, et al. Health care providers' experiences of pain management and attitudes towards digitally supported self-management interventions for chronic pain: a qualitative study. BMC Health Serv Res 2021;21(1):275 [FREE Full text] [doi: 10.1186/s12913-021-06278-7] [Medline: 33766028]

43. Boström K, Børøsund E, Varsi C, Eide H, Flakk Nordang E, Schreurs KM, et al. Digital self-management in support of patients living with chronic pain: feasibility pilot study. JMIR Form Res 2020;4(10):e23893 [FREE Full text] [doi: 10.2196/23893] [Medline: 33094734]

44. Ledel Solem IK, Varsi C, Eide H, Kristjansdottir OB, Børøsund E, Schreurs KM, et al. A user-centered approach to an evidence-based electronic health pain management intervention for people with chronic pain: design and development of EPIO. J Med Internet Res 2020;22(1):e15889 [FREE Full text] [doi: 10.2196/15889] [Medline: 31961331]

45. Luborsky MR, Rubinstein RL. Sampling in qualitative research: rationale, issues, and methods. Res Aging 1995;17(1):89-113 [FREE Full text] [doi: 10.1177/0164027595171005] [Medline: 22058580]

46. Thornquist E, Bunkan BH. Hva er psykomotorisk behandling?. Oslo: Universitetsforlaget; 1986.

47. Bunkan BH. Kropp, respirasjon og kroppsbilde: ressursorientert kroppsundersøkelse og behandling. Oslo: Universitetsforlaget; 1996.

48. Malterud K. Systematic text condensation: a strategy for qualitative analysis. Scand J Public Health 2012;40(8):795-805. [doi: 10.1177/1403494812465030] [Medline: 23221918]

49. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]

50. Jennings EM, Okine BN, Roche M, Finn DP. Stress-induced hyperalgesia. Prog Neurobiol 2014;121:1-18. [doi: 10.1016/j.pneurobio.2014.06.003] [Medline: 25010858]

51. Vachon-Presseau E, Roy M, Martel MO, Caron E, Marin MF, Chen J, et al. The stress model of chronic pain: evidence from basal cortisol and hippocampal structure and function in humans. Brain 2013;136(Pt 3):815-827. [doi: 10.1093/brain/aws371] [Medline: 23436504]

52. Lumley MA, Sklar ER, Carty JN. Emotional disclosure interventions for chronic pain: from the laboratory to the clinic. Transl Behav Med 2012;2(1):73-81 [FREE Full text] [doi: 10.1007/s13142-011-0085-4] [Medline: 22905067]

53. Serbic D, Pincus T. The relationship between pain, disability, guilt and acceptance in low back pain: a mediation analysis. J Behav Med 2017;40(4):651-658 [FREE Full text] [doi: 10.1007/s10865-017-9826-2] [Medline: 28155002]

54. Vlaeyen JW, Kole-Snijders AM, Boeren RG, van Eek H. Fear of movement/(re)injury in chronic low back pain and its relation to behavioral performance. Pain 1995;62(3):363-372. [doi: 10.1016/0304-3959(94)00279-N] [Medline: 8657437]

55. Vlaeyen JW, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. Pain 2000;85(3):317-332. [doi: 10.1016/S0304-3959(99)00242-0] [Medline: 10781906]

56. Jafari H, Courtois I, Van den Bergh O, Vlaeyen JW, Van Diest I. Pain and respiration: a systematic review. Pain 2017;158(6):995-1006. [doi: 10.1097/j.pain.0000000000000865] [Medline: 28240995]

57. Abell JE, Hootman JM, Zack MM, Moriarty D, Helmick CG. Physical activity and health related quality of life among people with arthritis. J Epidemiol Community Health 2005;59(5):380-385. [doi: 10.1136/jech.2004.028068] [Medline: 15831686]

58. Sivertsen B, Lallukka T, Petrie KJ, Steingrímsdóttir Ó, Stubhaug A, Nielsen CS. Sleep and pain sensitivity in adults. Pain 2015;156(8):1433-1439. [doi: 10.1097/j.pain.0000000000000131] [Medline: 25915149]

59. Bandura A. Self-efficacy: the exercise of control. New York: W H Freeman/Times Books/ Henry Holt & Co; 1997.

60. Muraven M, Tice DM, Baumeister RF. Self-control as limited resource: regulatory depletion patterns. J Pers Soc Psychol 1998;74(3):774-789. [doi: 10.1037//0022-3514.74.3.774] [Medline: 9523419]

61. McDermott MS, While AE. Maximizing the healthcare environment: a systematic review exploring the potential of computer technology to promote self-management of chronic illness in healthcare settings. Patient Educ Couns 2013;92(1):13-22. [doi: 10.1016/j.pec.2013.02.014] [Medline: 23566427]

62. Grencavage LM, Norcross JC. Where are the commonalities among the therapeutic common factors? Prof Psychol Res Pract 1990;21(5):372-378. [doi: 10.1037/0735-7028.21.5.372]

63. Weinberger J. Common factors aren't so common: the common factors dilemma. Clin Psychol Sci Pract 1995;2(1):45-69. [doi: 10.1111/j.1468-2850.1995.tb00024.x]

64. Korstjens I, Moser A. Series: practical guidance to qualitative research. Part 4: trustworthiness and publishing. Eur J Gen Pract 2018;24(1):120-124. [doi: 10.1080/13814788.2017.1375092] [Medline: 29202616]

Abbreviations

ACT: acceptance and commitment therapy
CBT: cognitive behavioral therapy
PMP: psychomotor physiotherapy

XSL•FO
RenderX

Original Paper

Culturally Tailored Social Media Content to Reach Latinx Immigrant Sexual Minority Men for HIV Prevention: Web-Based Feasibility Study

Jane J Lee[1], MSW, PhD; Joel Aguirre Herrera[2], BA; José Cardona[2], BA; Loren Yesenia Cruz[2]; Lésster Munguía[2], BS; Christopher A Leyva Vera[1], BASW, MSW; Gabriel Robles[3], LCSW, PhD

[1]School of Social Work, University of Washington, Seattle, WA, United States
[2]Entre Hermanos, Seattle, WA, United States
[3]School of Social Work, Rutgers University, New Brunswick, NJ, United States

Corresponding Author:
Jane J Lee, MSW, PhD
School of Social Work
University of Washington
4101 15th Avenue NE
Seattle, WA, 98105
United States
Phone: 1 2066168984
Email: janejlee@uw.edu

Abstract

Background: Latinx gay, bisexual, and other sexual minority men are disproportionately affected by HIV in the United States. As Latinx sexual minority men, particularly those who are foreign-born, experience inequitable access to health services, tailored strategies to engage them for HIV prevention are urgently needed.

Objective: Our study seeks to address the need for enhanced access to HIV prevention among Latinx immigrant sexual minority men. We developed and piloted a culturally sensitive technology-based campaign focused on HIV testing and pre-exposure prophylaxis (PrEP) uptake.

Methods: We used a two-phase approach to assess the feasibility of community-informed social media content in engaging Latinx immigrant sexual minority men for HIV testing and PrEP use. First, we conducted three iterative focus groups with 15 Latinx immigrant sexual minority men to refine the HIV prevention content to be piloted on social media platforms. The finalized content was placed on Instagram and Facebook for 9 days in July and September 2021 to individuals who were in Washington State. Individuals who clicked on the content were directed to a website with additional HIV prevention information. Second, we conducted online surveys (n=60) with website visitors that assessed sociodemographic characteristics, barriers to HIV prevention, and HIV-related transmission risk and prevention behaviors. We conducted descriptive analyses to examine the overall profile of survey respondents and determine the feasibility of culturally informed social media content in reaching Latinx immigrant sexual minority men.

Results: Overall, 739 unique users visited the website during the 9-day period when the social media content was posted on Instagram and Facebook. Our sample included 60 Latinx immigrant sexual minority men who completed the online survey. Participants' mean age was 30.8 years and more than half (n=34, 57%) completed the survey in Spanish. A quarter of participants indicated that they were unauthorized immigrants and 57% (n=34) reported not having medical insurance. Participants reported, on average, having 6 different sexual partners in the last 6 months. Nearly a third of respondents had not tested for HIV in the last 6 months. Only about half (n=32, 53%) of respondents had used PrEP in the last 12 months.

Conclusions: Community-driven social media and web-based strategies are feasible ways to engage Latinx immigrant sexual minority men who may traditionally lack access to HIV prevention information and services due to structural and social barriers. The results highlight that culturally relevant social media and web-based outreach strategies that are informed and developed by the community can reach Latinx immigrant sexual minority men for HIV prevention. Findings underscore the need to examine the effectiveness of social media content in promoting HIV testing and PrEP uptake in marginalized Latinx populations.

XSL·FO
RenderX

Introduction

Despite significant progress in reducing the number of new HIV infections, there were more than 36,000 new HIV diagnoses in the United States and dependent areas in 2019 [1]. The vast majority of these new diagnoses occurred in gay, bisexual, and other sexual minority men, with Latinx sexual minority men accounting for nearly a quarter of all new HIV infections [1]. Latinx sexual minority men encounter significant social and structural challenges to seeking and receiving appropriate HIV prevention services [2,3]. Specifically, factors such as language barriers, immigration status, education level, and discrimination may be linked to lower uptake of HIV prevention strategies such as HIV testing and pre-exposure prophylaxis (PrEP) use in Latinx immigrant sexual minority men [2,3]. Further, with the unprecedented disruptions related to the COVID-19 pandemic, Latinx immigrant sexual minority men may be less likely to seek in-person care and information for nonurgent conditions, potentially delaying HIV testing and PrEP use. Hence, while effective strategies for HIV prevention and treatment are available, increased efforts to enhance their access and use among Latinx immigrant sexual minority men are urgently needed.

The use of eHealth and related technology has the potential to improve the reach and implementation of HIV prevention strategies for marginalized populations such as Latinx immigrant sexual minority men [4,5]. HIV prevention information delivered via online and social media platforms may offer greater privacy and confidentiality, which may increase the acceptability of accessing HIV prevention resources [5]. eHealth tools have also served as a useful mechanism to deliver information and health services when in-person activities are limited or unavailable due to public health concerns such as COVID-19 [6]. Further, the wide use and availability of the internet and social media applications have made eHealth HIV prevention approaches increasingly accessible for hard-to-reach populations [7,8]. For Latinx immigrant sexual minority men who may be reluctant about seeking health services due to stigma or fear, eHealth may address barriers and support engagement with HIV prevention resources. Yet, despite the growing number of eHealth HIV prevention interventions, there has been limited attention to the acceptability of using eHealth strategies with Latinx immigrant sexual minority men [9,10].

To address the need to tailor HIV prevention programs for specific groups, we sought to enhance understanding of the specific preferences and acceptability of using eHealth strategies for HIV prevention among Latinx immigrant sexual minority men. We conducted an initial qualitative study to examine how to appropriately use social media platforms to recruit and engage Latinx immigrant sexual minority men for HIV testing and PrEP uptake [11]. We also assessed how specific content delivered on social media platforms can address barriers to HIV testing and PrEP in this population. The findings of this study indicated that Latinx immigrant sexual minority men were enthusiastic about receiving HIV prevention information via social media platforms. Participants offered specific suggestions and preferences regarding how social media content should be developed and emphasized that messaging be inclusive, motivational, and positive [11].

Based on the results of this initial study, which are published elsewhere [11], we developed content to be piloted on social media platforms to engage Latinx immigrant sexual minority men for HIV testing and PrEP uptake. This study examines the feasibility of using the culturally tailored content on social media platforms to reach Latinx immigrant sexual minority men for HIV prevention.

Methods

Overview

This feasibility study used a community-based participatory approach and was conducted in Seattle, Washington in partnership with a local community-based organization that serves the lesbian, gay, bisexual, transgender, and queer (LGBTQ+) Latinx community. The first phase of the study involved three focus groups that were developed to finalize the social media content for the study. During the study's second phase, the finalized content was piloted on social media platforms, and a web-based survey was conducted to assess whether the social media content was successful in reaching Latinx immigrant sexual minority men for HIV testing and PrEP uptake.

Focus Groups

We developed five initial designs of the culturally tailored HIV prevention content to be piloted on social media platforms based on the preferences identified from Latinx immigrant sexual minority men. We conducted three iterative focus groups with 15 Latinx immigrant sexual minority men to further revise the initial designs according to participants' preferences. During each focus group, participants offered suggestions to edit the content and ensure the cultural sensitivity of the designs. Participants also discussed the rationale behind their suggested changes. We finalized the designs to be piloted on social media platforms after the third and final focus group.

Pilot Testing of the Social Media Content for Latinx Immigrant Sexual Minority Men

The social media content was piloted on social media sites in July 2021. Specifically, the content was placed on Facebook and Instagram over a 9-day period. The content was available to individuals who were on Instagram or Facebook in Washington State. Individuals who clicked on the content were directed to a website that offered additional information on HIV testing and PrEP. Specifically, the website was developed with support from the study's community partner and offered

localized and culturally relevant HIV prevention information in Spanish. Information about how and why to get tested for HIV and use PrEP were included. Additionally, contact information to make an appointment to test for HIV or learn more about PrEP were also available on the website. We conducted an online survey with 60 website visitors to assess whether the social media content was a feasible strategy for reaching Latinx immigrant sexual minority men for HIV prevention. To be eligible to complete the survey, individuals had to meet the following inclusion criteria: be 18 years or older, identify as Hispanic or Latino/a/x, male sex at birth, born outside of the continental United States, and report sex with men. All individuals consented to participate prior to completing the survey. Surveys were completed in Spanish or English based on the participant's preference.

Measures

Surveys were available in Spanish or English and assessed sociodemographic characteristics, barriers to HIV prevention, and HIV risk and prevention behaviors. Barriers to HIV prevention included not having health insurance and sexual orientation disclosure (whether or not individuals disclosed their sexual orientation to friends who are heterosexual, family members, employers or teachers, and health care providers). We also assessed participants' perceptions of their community's tolerance toward individuals who identify as LGBTQ+ (1 strongly disagree, 2 disagree, 3 neither agree nor disagree, 4 agree, 5 strongly agree). Surveys also measured the extent to which participants experienced distress related to COVID-19 (1 no distress, 10 extreme distress). HIV transmission risk behaviors were assessed by the number of sexual partners in the last 6 months, drug or alcohol use before last sex, and condomless anal sex at last sex. HIV prevention behaviors included ever testing for HIV, testing for HIV in the last 6 months, PrEP awareness, and PrEP use in the last 12 months. Descriptive analyses were used to examine the overall profile of website visitors and to determine the feasibility of using community-informed social media content to reach Latinx immigrant sexual minority men in need of HIV testing and PrEP who encounter barriers to HIV prevention.

Ethics Consideration

The authors obtained informed consent from all study participants prior to their participation in the focus groups or web-based surveys. All study procedures were reviewed by the University of Washington Human Subjects Division and qualified for exempt status from federal human subjects regulations.

Results

Focus Groups

The focus groups provided feedback on the language, imagery, and overall designs of the community-informed social media content. The initial versions of the content were revised in the following ways based on focus group feedback: photographs were changed to depict more realistic images of Latinx sexual minority men, additional colors and illustrations were integrated to highlight the empowering aspects of HIV prevention, and designs that were text-focused were included to provide more discreet ways to engage with the content. The finalized content is presented in Figures 1-5.

Figure 1. Culturally tailored social media content for HIV prevention in Latinx immigrant sexual minority men: Be your hero.

Figure 2. Culturally tailored social media content for HIV prevention in Latinx immigrant sexual minority men: Illustrated men. PrEP: pre-exposure prophylaxis.

Figure 3. Culturally tailored social media content for HIV prevention in Latinx immigrant sexual minority men: Favorite sexual position. PrEP: pre-exposure prophylaxis.
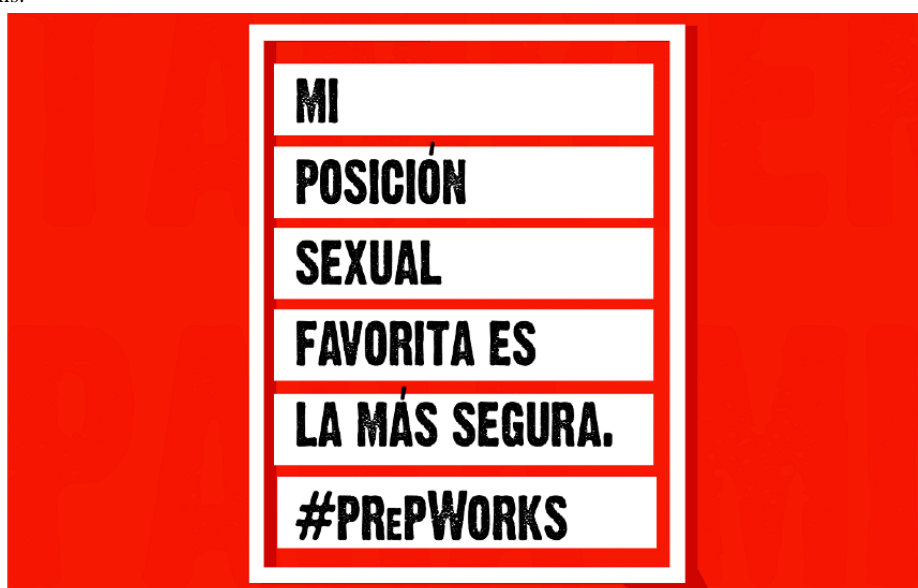
Figure 4. Culturally tailored social media content for HIV prevention in Latinx immigrant sexual minority men: For love. PrEP: pre-exposure prophylaxis.

Figure 5. Culturally tailored social media content for HIV prevention in Latinx immigrant sexual minority men: Emojis. PrEP: pre-exposure prophylaxis.

Figure 1 states "Take charge of your life with PrEP" and "Be your hero" in Spanish, and presents a multimedia image of a man in the sky. Figure 2 presents an illustrated image of two men and a symbol of PrEP. Figure 3 is text-focused and states, "My favorite sexual position is the one that's safest" in Spanish with the hashtag "#PrEPWorks." Figure 4 includes an image of two men with the words "For love," "For me," "For him," in Spanish and "PrEP pill." Figure 5 states, "It's okay to play with your food" in Spanish and uses emojis and illustrations to demonstrate that sex with condoms and with PrEP is the "safest sex." These five designs reflect community priorities, preferences, and needs to receiving information and services related to HIV prevention. For example, the use of "Spanglish" or both English and Spanish in the designs was encouraged as it highlights that many Latinx immigrant sexual minority men speak both Spanish and English. Additionally, participants noted that the use of humor and cheeky language and imagery were an important way to catch their attention. Overall, the content relays the positive and motivational messages sought by focus group participants.

Characteristics of Feasibility Study Survey Respondents

There were 739 unique users who visited the website after clicking the pilot social media content on Facebook or Instagram over the 9-day period. Among the 60 Latinx immigrant sexual minority men who completed the online survey, slightly more than half (n=34, 57%) completed the survey in Spanish. Respondents' sociodemographic characteristics are presented in Table 1.

The mean age was 31 years, and the average length of time residing in the United States was 14 years. About half of the respondents (n=32, 53%) were born in Mexico and 20% (n=12) reported having less than a high school level education. More than half of the participants (n=35, 58%) reported their religion as Catholic or Christian. A quarter of participants were unauthorized immigrants, while 17% (n=10) were legal permanent residents (Table 1).

Table 1. Characteristics of 60 Latinx immigrant sexual minority men who were reached by the culturally informed pilot social media content.

	Participants, n (%)
Age (years), mean (SD)	30.8 (8.2)
Country of origin, n (%)	
Mexico	32 (53)
El Salvador	6 (10)
Honduras	4 (7)
Cuba	4 (7)
Puerto Rico	4 (7)
Other[a]	10 (17)
Length of time in United States (years), mean (SD)	13.9 (8.5)
Education, n (%)	
Less than high school	12 (20)
High school or GED[b]	14 (23)
Some college	13 (22)
Bachelor's degree or higher	21 (35)
Employment status, n (%)	
Unemployed	9 (15)
Employed full time	32 (53)
Employed part time	14 (23)
In school	4 (7)
Retired	1 (2)
Religion, n (%)	
Catholic	18 (30)
Protestant/Christian	17 (28)
No religion	11 (18)
Other[c]	8 (13)
Atheist	4 (7)
Immigration status, n (%)	
Legal permanent resident	10 (17)
Naturalized citizen	8 (13)
Unauthorized immigrant	15 (25)
Eligible immigrant	24 (40)
Temporary resident or other	3 (5)

[a]Other included Argentina, Brazil, Panama, Guatemala, Colombia, Ecuador, and Peru.

[b]GED: General Educational Development.

[c]Other included: Nondenominational or independent, spiritual but not religious, and other.

Barriers to HIV Prevention

Select barriers to HIV prevention are presented in Table 2.

Approximately 42% (n=25) of respondents indicated that they did not have health insurance. While the majority (n=53, 88%) of participants reported disclosing their sexual orientation to their health care providers, less than three-fourths had disclosed their sexual orientation to friends who are heterosexual or to family members. About half (n=31, 52%) of respondents had disclosed their sexual orientation to employers or teachers. Participants reported an average rating of 3.5 (range 1-5) in community tolerance toward LGBTQ+ individuals. On a scale of 1 to 10 with "1" being no distress and "10" being extreme distress, respondents reported experiencing an average level of 5.0 in COVID-19–related distress.

XSL•FO
RenderX

Table 2. Barriers to HIV prevention and HIV transmission risk and prevention behaviors among Latinx immigrant sexual minority men (N=60).

	Participants, n (%)
Health insurance, n (%)	
Yes	34 (57)
No	25 (42)
Unsure	1 (2)
COVID-19–related distress, mean (SD)	5.0 (2.9)
Community LGBTQ+[a] tolerance, mean (SD)	3.5 (1.3)
Sexual orientation disclosure[b], n (%)	
Disclosed to friends who are heterosexual	43 (72)
Disclosed to family members	42 (70)
Employers or teachers	31 (52)
Disclosed to health care providers	53 (88)
Number of sexual partners in last 6 months, mean (SD)	6.2 (7.1)
Drug or alcohol use behavior during last sex[c], n (%)	18 (30)
Condomless anal sex at last sex[c], n (%)	24 (40)
Ever tested for HIV[c], n (%)	48 (80)
Tested for HIV in last 6 months[c], n (%)	42 (70)
Currently using PrEP[c,d], n (%)	28 (47)

[a]LGBTQ+: lesbian, gay, bisexual, transgender, and queer.

[b]Data presented indicates respondents who disclosed.

[c]Data presented indicates respondents who reported "Yes" to the question.

[d]PrEP: pre-exposure prophylaxis.

HIV Transmission Risk and Prevention Behaviors

Table 2 also presents respondents' HIV transmission risk and prevention behaviors. Respondents reported an average of 6.2 sexual partners in the last 6 months and 30% (n=18) indicated using alcohol or drugs during last sex. While the majority (n=48, 80%) of participants had ever tested for HIV, 70% (n=42) had tested for HIV in the last 6 months. Less than half (n=28, 47%) of all respondents were currently using PrEP (Table 2).

Discussion

Primary Findings

Our results demonstrate that the pilot social media content reached a diverse sample of Latinx immigrant sexual minority men. Specifically, we engaged Latinx immigrant sexual minority men who were undocumented, had low levels of education, and were unemployed. While these factors often present challenges to accessing HIV prevention services [3,12,13], culturally tailored social media outreach strategies may provide unique opportunities to obtain HIV prevention information and resources that are traditionally out of reach for these communities. Further, more than half of respondents identified as Catholic or Protestant/Christian, which suggests that our social media content was a feasible approach to engaging Latinx sexual minority men who are members of religious organizations. Prior research has documented that Latinx sexual minority men who are religiously affiliated may experience greater internalized homophobia, which can have detrimental effects on HIV prevention and overall health [14]. While additional studies are needed to better understand the role of religion on HIV prevention in Latinx sexual minority men [15], culturally tailored HIV prevention content delivered via social media applications may provide a potential way to engage religious Latinx communities who may be deterred to access such information due to perceptions of internalized or societal stigma.

Our pilot content was also successful in reaching Latinx immigrant sexual minority men who reported several barriers to HIV prevention, including being uninsured. Not having health insurance can lead to missed opportunities for HIV testing and limit access to preventive health services [16]. As Latinxs have the highest uninsured rate of any racial or ethnic group in the United States [17], eHealth may be a valuable tool for reaching individuals who experience systemic health inequities. eHealth may also be a feasible HIV prevention strategy among Latinx immigrant sexual minority men in diverse contexts. Despite overwhelming challenges due to the COVID-19 pandemic, Latinx immigrant sexual minority men who reported experiencing distress related to COVID-19 engaged with the tailored HIV prevention social media content. Additionally, many respondents reported feeling that their communities were somewhat intolerant of individuals who identify as LGBTQ+.

As sexual minority men who face intolerance and discrimination may be less likely to use HIV prevention services, prevention services and information available through social media platforms may facilitate use among Latinx immigrant sexual minority men by creating a more acceptable environment in which to engage.

Further, a sizeable percent of respondents noted that they had not disclosed their sexual orientation to others, suggesting that many Latinx immigrant sexual minority men who were reached by the social media content may have concerns about prejudice and discrimination in their various social networks. Social media platforms and other eHealth tools can offer methods that may be more acceptable for individuals who seek to conceal their sexual orientation from specific groups and desire a discreet way to obtain HIV prevention services. Taken together, the tailored social media content was a feasible way to reach Latinx immigrant sexual minority men for HIV prevention given that its format and delivery method may overcome prior barriers to using services.

Notably, a considerable percentage of the Latinx immigrant sexual minority men reported engaging in sexual risk behaviors, including having multiple sexual partners, having condomless anal sex, and using drugs or alcohol before sex. Our results also demonstrated gaps in HIV prevention with nearly a third of respondents not having tested for HIV in the last 6 months and more than half not currently using PrEP. Hence, culturally tailored social media content is a feasible way to reach Latinx immigrant sexual minority men who not only experience barriers to HIV prevention but who also exhibit behaviors that may place them at elevated risk for HIV transmission.

Limitations

There are several limitations to this study. First, given that this study was designed to pilot the culturally informed content on Instagram and Facebook, a comparison to Latinx immigrant sexual minority men who did not engage with the social media content was not possible. Despite not having a comparison group, the authors believe that the characteristics of survey respondents highlight the opportunity that eHealth presents for reaching Latinx immigrant sexual minority men in need of HIV prevention services. Second, while the social media content was delivered virtually, the platforms focused on reaching individuals in Washington State and its Latinx population. Therefore, the social media content may not be appropriate or feasible for reaching all Latinx immigrant sexual minority men in the United States, including other locales with different compositions of Latinx immigrant populations. Further, our survey did not capture all individuals who engaged with or were impacted by the social media content. Hence, individuals who did not complete the survey may vary from those reported in this study.

Conclusions

Our culturally relevant social media and web-based outreach strategies that were informed and developed by the community are a feasible way to reach Latinx immigrant sexual minority men for HIV prevention. Community-driven social media and web-based content present opportunities for engaging individuals who may traditionally lack access to HIV prevention information and services. Future research may examine the effectiveness of culturally informed social media content in increasing HIV testing and PrEP use among Latinx immigrant sexual minority men.

Conflicts of Interest

None declared.

References

1.   Centers for Disease Control and Prevention. Estimated HIV incidence and prevalence in the United States, 2015-2019. HIV Surveillance Supplemental Report 2021 May;26:1-81.
2.   Sullivan PS, Satcher Johnson A, Pembleton ES, Stephenson R, Justice AC, Althoff KN, et al. Epidemiology of HIV in the USA: epidemic burden, inequities, contexts, and responses. Lancet 2021 Mar 20;397(10279):1095-1106. [doi: 10.1016/S0140-6736(21)00395-0] [Medline: 33617774]
3.   HIV and Hispanic/Latino gay and bisexual men. Centers for Disease Control and Prevention. 2019. URL: https://www.cdc.gov/hiv/pdf/group/gay-bisexual-men/hispanic-latino/cdc-hiv-group-gay-bisexual-men-hispanic-latino-factsheet.pdf [accessed 2022-03-08]
4.   Schnall R, Travers J, Rojas M, Carballo-Diéguez A. eHealth interventions for HIV prevention in high-risk men who have sex with men: a systematic review. J Med Internet Res 2014 May 26;16(5):e134 [FREE Full text] [doi: 10.2196/jmir.3393] [Medline: 24862459]

XSL•FO

RenderX

5.   Nguyen LH, Tran BX, Rocha LEC, Nguyen HLT, Yang C, Latkin CA, et al. A systematic review of eHealth interventions addressing HIV/STI prevention among men who have sex with men. AIDS Behav 2019 Sep;23(9):2253-2272 [FREE Full text] [doi: 10.1007/s10461-019-02626-1] [Medline: 31401741]

6.   Bokolo AJ. Application of telemedicine and eHealth technology for clinical services in response to COVID-19 pandemic. Health Technol (Berl) 2021;11(2):359-366 [FREE Full text] [doi: 10.1007/s12553-020-00516-4] [Medline: 33469474]

7.   Schueller SM, Muñoz RF, Mohr DC. Realizing the potential of behavioral intervention technologies. Curr Dir Psychol Sci 2013 Dec 03;22(6):478-483. [doi: 10.1177/0963721413495872]

8.   Li DH, Brown CH, Gallo C, Morgan E, Sullivan PS, Young SD, et al. Design considerations for implementing eHealth behavioral interventions for HIV prevention in evolving sociotechnical landscapes. Curr HIV/AIDS Rep 2019 Aug;16(4):335-348 [FREE Full text] [doi: 10.1007/s11904-019-00455-4] [Medline: 31250195]

9.   Solorio R, Norton-Shelpuk P, Forehand M, Martinez M, Aguirre J. HIV prevention messages targeting young Latino immigrant MSM. AIDS Res Treat 2014;2014:353092. [doi: 10.1155/2014/353092] [Medline: 24864201]

10. Martinez O, Wu E, Shultz AZ, Capote J, López Rios J, Sandfort T, et al. Still a hard-to-reach population? Using social media to recruit Latino gay couples for an HIV intervention adaptation study. J Med Internet Res 2014 Apr 24;16(4):e113 [FREE Full text] [doi: 10.2196/jmir.3311] [Medline: 24763130]

11. Lee JJ, Aguirre J, Munguia L, Robles G, Ramirez Hernandez K, Ramirez JI, et al. Engagement of Latino immigrant men who have sex with men for HIV prevention through eHealth: preferences across social media platforms. Ethn Health 2021 Jun 21:1-14. [doi: 10.1080/13557858.2021.1943322] [Medline: 34152248]

12. Oster AM, Russell K, Wiegand RE, Valverde E, Forrest DW, Cribbin M, NHBS Study Group. HIV infection and testing among Latino men who have sex with men in the United States: the role of location of birth and other social determinants. PLoS One 2013;8(9):e73779 [FREE Full text] [doi: 10.1371/journal.pone.0073779] [Medline: 24147151]

13. Brooks RA, Landrian A, Lazalde G, Galvan FH, Liu H, Chen Y. Predictors of awareness, accessibility and acceptability of pre-exposure prophylaxis (PrEP) among English- and Spanish-speaking Latino men who have sex with men in Los Angeles, California. J Immigr Minor Health 2020 Aug;22(4):708-716. [doi: 10.1007/s10903-019-00955-w] [Medline: 31823164]

14. Barnes DM, Meyer IH. Religious affiliation, internalized homophobia, and mental health in lesbians, gay men, and bisexuals. Am J Orthopsychiatry 2012 Oct;82(4):505-515 [FREE Full text] [doi: 10.1111/j.1939-0025.2012.01185.x] [Medline: 23039348]

15. Drumhiller K, Nanín JE, Gaul Z, Sutton MY. The influence of religion and spirituality on HIV prevention among Black and Latino men who have sex with men, New York City. J Relig Health 2018 Oct;57(5):1931-1947. [doi: 10.1007/s10943-018-0626-y] [Medline: 29696488]

16. Derose KP, Escarce JJ, Lurie N. Immigrants and health care: sources of vulnerability. Health Aff (Millwood) 2007;26(5):1258-1268. [doi: 10.1377/hlthaff.26.5.1258] [Medline: 17848435]

17. Keisler-Starkey K, Bunch LN. Health Insurance Coverage in the United States: 2019. US Census Bureau. Washington, DC: US Government Publishing Office; 2020 Sep. URL: https://www.census.gov/content/dam/Census/library/publications/2020/demo/p60-271.pdf [accessed 2022-02-08]

18. Shamah N. Cake Creative Agency. URL: https://cakeatwork.com/ [accessed 2022-03-12]

Abbreviations

LGBTQ+: lesbian, gay, bisexual, transgender, and queer
PrEP: pre-exposure prophylaxis

XSL·FO

RenderX

Media Source Characteristics Regarding Food Fraud Misinformation According to the Health Information National Trends Survey (HINTS) in China: Comparative Study

Angela Chang[1,2*], PhD; Peter Johannes Schulz[2*], PhD; Wen Jiao[1], MA; Guoming Yu[3], PhD; Ya Yang[3], PhD

[1]Faculty of Social Sciences, University of Macau, Macao, China

[2]Institute of Communication and Health, Lugano University, Lugano, Switzerland

[3]School of Journalism and Communication, Beijing Normal University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Ya Yang, PhD
School of Journalism and Communication
Beijing Normal University
#15 Xinjiekouwai Street
Haidian District
Beijing, 100875
China
Phone: 86 18810305219
Email: yangya@bnu.edu.cn

Abstract

Background: Ongoing rumors and fake news regarding food fraud, adulteration, and contamination are highly visible. Health risk information circulating through media and interpersonal communication channels has made *health crisis* an important research agenda.

Objective: This study explored the issue of food fraud and the effect of misinformation. Further, it assessed whether and how these issues have provided evidence-based interventions for food handlers and regulators to mitigate fraud misinformation.

Methods: The Health Information National Trends Survey (HINTS) was adopted for a collaborative study in China, after which a cross-sectional survey with door-to-door interviews was performed. Participants from Beijing and Hefei were selected using multistage sampling of adults in May 2017. Based on 4 government surveillance reports on food rumors and safety incidents, a descriptive analysis, correlation analysis, and analysis of variance were performed on the data.

Results: A total of 3090 results were gathered and analyzed. Among the respondents, 83.6% (2584/3090) heard at least one food rumor. Learning about food fraud was correlated with interpersonal connections (eg, doctors or health specialists) for accessing food health information. Overall, Chinese citizens with a higher level of interpersonal connection were more likely to be concerned about food incidents with a statistical difference (*P*<.001). Interpersonal connection was the most frequent communication source (698/1253, 55.7%), followed by traditional media (325/1253, 25.9%) and internet portals (144/1253, 11.5%). There was a significant relationship between media use and media category in Beijing (*P*<.001) and Hefei (*P*<.001). Overall, responses to food fraud and incident risks were lower in Beijing than in Hefei (*P*=.006). The respondents in Beijing were confronted more frequently by food rumors (range 346-1253) than those in Hefei (range 155-946). The urban dwellers in Beijing and their rural counterparts in Hefei also differed in terms of perceiving different levels of food risks from different media sources. The food rumor narratives that examined the conspiracy belief showed that social media played more important roles in influencing attitudes against misinformation for users in Hefei than in Beijing.

Conclusions: This study shows that consumers have to be on guard against not only fake food, but also spreading fake information and rumors, as well as conspiracy beliefs involving fake food. This study focused on characterizing media sources, types of food fraud misinformation, and risk perceptions of food safety, which mix urgency and suspicion, and attempted to provide evidence-based interventions for risk management guidance, with the hypothesis of significant correlations between media types and sources, and consumer exposure and perception levels of food rumors and risks.

XSL•FO
RenderX

Introduction

Background

Food fraud has been practiced since ancient times but has become more sophisticated in recent years. The recent food supply chains have become complicated and accelerated in local markets. Government authorities may be able to trace some imported contaminated food items; however, they are smuggled from one country to another with switched packaging and altered production dates. Thus, the risk of food fraud has broadened to include the global population [1]. Food fraud that involves items, such as infant milk formula, olive oil, and Scotch whisky, has been examined for perception and attitude learning among consumers in first- and second-tier cities in China [2], and the conclusion indicated that food fraud represents a food hazard and poses a threat to the authenticity, quality, and reliability of food.

Investigating food fraud is an emerging research idea and is a newly defined area of food protection [1,3,4]. Food fraud is an intentional act for economic gain, whereas a food safety incident may be an unintentional act with unintentional harm [5]. Thus, studying food fraud and adulterants mainly focuses on food safety (eg, pesticide residue) and food defense with malicious intent to harm (eg, terrorism attack). Food fraud and adulteration are principally motivated by economic drive; however, food fraud–related public health risks are often riskier than traditional food safety threats [1,6]. This is because food contaminants are complicated risks, which may involve chemical, microbial, physical, and allergenic contaminants with food fraud conditions. This also explains why food handlers have a responsibility to ensure that the food they prepare is free from these contaminants and safe for the consumer [7].

In recent decades, health information has been particularly exposed to rumors, incidents, and fake news [6,8,9]. Until the recent 10 years, little was known about the access, sources, and trust levels of food fraud and related information, or factors that facilitate or hinder communication on a population-wide basis [10-12]. In order to help fill this gap, researchers have tried to employ conspiracy beliefs to explain health-related matters. Conspiracy beliefs adjoined to health information are reportedly on the rise and are beginning to receive significant attention (eg, water fluoridation, vaccines, cell phones, unhealthy diets, and alternative medicine) [9,13-15].

More recently, the outbreak of food incidents and rumors has provided evidence of how the spread of conspiracy thinking has reached beyond the narrow confines of individual or group believers [16-18]. A conspiracy belief works through emotional appeals and implausibility, which can be used as a rhetorical device to appeal to the emotions of a significant number of the public [19,20]. The conspiracy belief is significantly associated with health worries while considering variables such as demographics, ideology, and health perceptions [8,21]. However, it remains unclear whether the objective information of conspiracy beliefs matters.

China's food safety rumors and concerns involve fraud, implicated foods, adulterants, contaminants, and abnormal conditions, as well as concerned food sources [4]. For many in China, the term genetically modified foods (GMFs) evokes negative reactions as it is usually connected with food safety concerns such as poisoned seeds or contaminated fields [22]. Vast food fraud and related studies have focused on GMFs for supposed damage [23-26].

While academia researchers, government regulators, and industry practitioners alike struggle with this evolving issue [4,27], the terms and definitions of food fraud should include deliberate substitution, dilution, counterfeiting, or misrepresentation of food, ingredients, or packaging, as well as false or misleading statements made about a food product. All these examples of food fraud can have a negative impact on the quality and safety aspects of food, some of which have resulted in serious illnesses and even fatalities, damaging trust in domestically produced and imported food.

Media Exposure of Food Health Risk

Little is known about the sources and trust of food fraud information hindering communication on a population-wide basis. The Health Information National Trends Survey (HINTS) was developed to help fill this gap [28,29]. The HINTS reported that around 50% of American respondents preferred to go to their physicians for specific health information [30]; however, when asked where they actually went, up to 48.6% of respondents reported going online first.

People's perceptions of health risks are initially determined by the scope and width of coverage from the media for the most part [8]. With the growing influence of social media in the public sphere, the communication of alternative health discourse often circulates in opposition to the mainstream or government-owned media [31-33]. Consequentially, the narratives supporting alternative health discourses have increasingly become the growing consensus for laymen rather than health professionals or policymakers [24,34,35].

In order to determine the extent of health conspiracism in public, social media might play an increasingly important role in active information-seeking behavior [15,36]. Although previous studies provide important insights for understanding active seeking behavior, they neglect the incidental seeking behavior of health information.

The steady growth of food safety rumors and incidents would increase citizens' concerns, but media themselves inform, circulate, and even amplify the food safety risk perceived. In fact, tensions arise between medical science protecting the collective well-being and the emotive amplification of groups concerned with individual health. Such exchanges have developed in food health and safety discourses, with online forums working as echo chambers [37].

Goal of This Study

The narratives of scandals and rumors might correlate with media sources and impact actual health information seeking behaviors. Thus, this study focused on characterizing media sources, the types of food fraud misinformation, and the risk perception of food safety, which mixes urgency and suspicion.

The information regarding food risks and safety is becoming more relevant [7]. If information is viewed in a more comprehensive and complex way, the resulting importance of the topic would encourage clarification of the types of food fraud and consequentially which conspiracy beliefs are salient and whether the public has prescribed to these beliefs.

Research Questions

The abovementioned conspiracy beliefs are associated with the behavior of seeking information that impacts food fraud. Thus, the risk perceptions of food could be different for passive information seeking behavior. Concomitantly, demographic variables are significantly related to the concerns of food fraud and adulteration type [7,38]. Thus, the following 5 research questions were raised: (1) What types of food fraud–related incidents and rumors are spreading fear in China, regarding food safety concerns? (2) Are the conspiracy beliefs associated with demographic variables, such as income, age, gender, and education? (3) Is learning about food fraud correlated with media type and source? (4) Are citizens with more social media exposure more likely to be concerned about food rumors and scandals? (5) Is there any difference in the perception level of food risk between urban dwellers and their rural counterparts?

Methods

Data Collection

The Health Information National Trends Survey in China (HINTS-China) project extends the established HINTS from the United States [11,39-41]. The questionnaire was built upon the research framework of the HINTS by using identical core questions, as well as adding unique food fraud health misinformation and concerns, and communication characteristics for the Chinese public. It is a cooperative project supported by Beijing Normal University, George Mason University, the National Cancer Institute in the United States, the Chinese Food and Drug Administration, and the Chinese Health Media Group. The HINTS-China was approved by the institutional review board of the School of Journalism and Communication, Beijing Normal University in 2017.

Data were collected in 2017 from Beijing, the capital city of the People's Republic of China, and Hefei, a second-tier and capital city in Anhui Province. Considering the different developments in political and economic applications, this study compared the commonalities and differences of the 2 locations to learn the impacts of food rumors and incidents on consumers [11,39,42,43]. With respect to urban and rural respondents, it aimed to provide a picture of the information communication strategy and risk assessment.

A multistage stratified sampling method was employed, and the procedure included training interviewers for door-to-door interviews between March to May in 2017. The interviews took place from May 9 to 24, 2017. One member aged 18 years or above from each household was requested to take part. Respondents who took part in the survey gave their written consent [43].

Conspiracy Narratives of Food Misinformation

The spread of malicious or accidental rumors online, particularly regarding food in real-world emergencies, can have harmful effects on the society. Thus, 8 highly publicized narratives of food risks were officially identified as rumors before implementing the HINTS-China. The veracity of food rumors comprised a range of food additives, fraudulent activities, adulterant substances, GMFs, and food contamination. For example, suspected rumors of food narratives were as follows: "seaweed (nori) is made out of black plastic," "big and sweet strawberries are with excessive swelling agent," and "seedless grapes are covered with contraceptives." A list of food fraud narratives with English translations is displayed in Textbox S1 in Multimedia Appendix 1.

Health Risks of Food Fraud

Eleven heated food safety incidents in 2016 were exposed. These were included in the HINTS-China to measure consumers' awareness and perception of health risks. The veracity of negative food exposure comprised the food-origin story, food additives, adulterants, authenticity, crime, contamination, integrity, quality, and food security. For instance, the narratives regarding food safety were as follows: "seafood from radiated areas in Japan were smuggled into China," "more than 100 tons of expired milk powder from New Zealand was illegally repackaged and sold in Shanghai," "A Vietnam yogurt brand with three-no yogurt (low-quality product with no manufacturer name, no production site, and no production hygiene license code) entered the market in many cities in mainland China, and the product was seized by the China Food & Drug Administration (CFDA)," and "A food ordering platform was accused of partnering with an unlicensed restaurant with poor sanitary conditions by a government-owned television program." A list of health risks of food incidents with safety concerns is shown in Textbox S2 in Multimedia Appendix 1.

Detection and Verification Mechanism

Many food risk concerns and incidents in China were initially reported by central official surveillance and then by local media. The state-owned news media detected food fraud and adulteration in news content and also delivered information concerning food authenticity and integrity. The salient features of rumors and food safety were identified from the following 2 official sources: People's Daily Online (人民網) and Xinhuanet (新華網) [4,44]. People's Daily is supported by the Central Committee of the Chinese Communist Party, while the XinHua Press Agency is an official agency and is the largest news agency in the world.

Two more sources joined the verification mechanism of food rumors and incidents, namely, Chinese Health Daily News (中国健康报) and CNPharm.com (中国食品药品网). Online media provide professional and reliable information under the supervision of the National Health Commission. It has been

confirmed that all of the 8 popular rumors in Textbox S1 in Multimedia Appendix 1 are false, while the narratives of the 11 food scandals in Textbox S2 in Multimedia Appendix 1 reflect the facts in 2016, though 2 of them are not accurate.

Measurements

We measured the conspiracy narratives of food fraud, rumors, incidents, safety issues, characteristics of media sources, and scale descriptions. In particular, the measurements of the narratives of food rumors and incidents were to learn whether the media sources that people searched for on hearing about a statement were trustworthy.

The established measures included a dichotomous answer of yes/no for having heard about each food rumor or incident. It was followed by a 5-point Likert scale for the measurement of the degree of trust, ranging from 1 (completely trust) to 5 (completely distrust), with a higher score indicating a higher distrust of the food rumor and a lower score indicating a lower distrust of the food rumor. The socioeconomic information measured included gender, age, ethnic group, occupation, marital status, highest level of education, and monthly income.

Regarding the media source, the HINTS-China measured the media use behavior regarding seeking general information and health information. The first question focused on 12 media characteristics, either general purpose or health/medical-related information in the past 12 months. A dichotomous question checked people's active information seeking process regarding health/medicine. Moreover, a total of 25 media sources were provided to learn their impact on food rumors and incidents. In order to merge the media sources, the following 5 categories were established for analysis: (1) interpersonal media, (2) public organizations, (3) traditional media, (4) internet portals, and (5) social media (Multimedia Appendix 2). One media option was excluded from the analysis because it appeared that none or very few people had accessed it (less than 1%).

Descriptive and correlation analyses were performed using SPSS 24 (IBM Corp). For any accessed media that connected with food misinformation and food safety issues, analysis of variance (ANOVA) was employed, which determined whether the influence of media use mediated the conspiracy narratives of food rumors and incidents.

Results

Descriptive Analysis

As of May 2017, we conducted the survey and data collection, which was funded by CFDA and Chinese Health Media Group in January 2017, and approved by the institutional review board at the School of Journalism and Communication, Beijing Normal University in March 2017. After 16 dates of data collection (May 9 to 24), 3090 samples were collected, and the number of valid questionnaires was 2584, including 1462 in Beijing and 1122 in Hefei, after list-wise deletion of missing values. The preliminary description was made in August 2017, and further data analysis was performed in May 2020.

The response rate was 42.65% for Beijing and 57.35% for Hefei (out of 3090 participants). The dominant ethnic group was Han Chinese, ranging from 98.0% (1433/1462) in Beijing to 99.5% (1116/1122) in Hefei. Respondents' ages ranged from 18 to 60 years (mean 38.02 years, SD 11.16) in Beijing, but the respondents were younger on average in Hefei (mean 32.26 years, SD 11.23). In Beijing, respondents were predominantly female (920/1462, 62.9%), married or in cohabitation (1199/1462, 82.0%), and high school educated (553/1462, 37.8%). In comparison, Hefei's respondents were predominantly female (677/1122, 60.3%), married or in cohabitation (696/1122, 62.0%), and college educated with a bachelor's degree (351/1122, 31.3%). Table 1 displays the sociodemographic findings of respondents in Beijing and Hefei in 2017.

Table 1. Sociodemographic findings of respondents in Beijing and Hefei in 2017.

Characteristic	Beijing (N=1462), n (%)	Hefei (N=1122), n (%)
Gender		
Male	542 (37.1)	445 (39.7)
Female	920 (62.9)	677 (60.3)
Occupation		
General staff in the organization	342 (23.4)	82 (7.3)
Worker or self-employed	307 (21.0)	289 (25.8)
Business service staff	229 (15.7)	45 (4.0)
Retiree or unemployed	267 (18.3)	151 (13.5)
Government administrator	15 (1.0)	29 (2.6)
Professional technician	95 (6.5)	101 (9.0)
Enterprise administrator	88 (6.0)	116 (10.3)
Education sector	37 (2.5)	54 (4.8)
Student	31 (2.1)	207 (18.4)
Private entrepreneur	24 (1.6)	36 (3.2)
Agricultural laborer	27 (1.8)	12 (1.1)
Marital status		
Married or cohabiting	1199 (82.0)	696 (62.0)
Single, never married	243 (16.6)	399 (35.6)
Education level		
Primary school or below	28 (1.9)	30 (2.7)
Junior middle school	288 (19.7)	148 (13.2)
High school	553 (37.8)	200 (17.8)
Junior college	362 (24.8)	308 (27.5)
Bachelor's degree	219 (15.0)	351 (31.3)
Master's degree or above	12 (0.8)	85 (7.6)
Individual income (RMB)[a]		
No income	76 (5.2)	243 (21.7)
Below 1000	11 (0.8)	28 (2.5)
1000-2499	194 (13.3)	203 (18.1)
2500-4999	719 (49.2)	378 (33.7)
5000-9999	413 (28.2)	210 (18.7)
10,000 or above	49 (3.4)	60 (5.3)

[a]1 RMB = 0.158 USD.

Overall, interpersonal communication was observed to be the most favorable first choice of media use for accessing food risk information (n=1466), followed by internet portals (n=464) and traditional media (n=425). A chi-square test of independence showed that there was a significant relationship between media use and media category in Beijing (χ^2_8=68.223; P<.001) and in Hefei (χ^2_8=78.529; P<.001). The rank of media preference greatly impacts food information outcomes. The additional insight shows that merging use of media sources mimics the way we now consume content across devices and platforms. Table 2 outlines the ranking of media sources for accessing food information in Beijing and Hefei.

Table 2. Ranking of media sources for accessing food risk information in Beijing and Hefei.

Media source	Choice of media in Beijing				Choice of media in Hefei			
	1st (n=572), n (%)	2nd (n=571), n (%)	3rd (n=572), n (%)	Subtotal, n	1st (n=426), n (%)	2nd (n=373), n (%)	3rd (n=410), n (%)	Subtotal, n
Interpersonal communication	385 (39.3)	331 (33.8)	263 (26.9)	979	209 (42.9)	155 (31.8)	123 (25.3)	487
Internet portals	69 (29.7)	65 (28.0)	98 (42.2)	232	98 (42.2)	72 (31.0)	62 (26.7)	232
Traditional media	75 (25.5)	111 (37.8)	108 (36.7)	294	34 (26.0)	24 (18.3)	73 (55.7)	131
Public organizations	34 (22.7)	47 (31.3)	69 (46.0)	150	63 (24.7)	90 (35.3)	102 (40.0)	255
Social media	9 (15.0)	17 (28.3)	34 (56.7)	60	22 (21.2)	32 (30.8)	50 (48.1)	104

Specifically, in Beijing, seeking doctors' or health specialists' advice concerning food risk information was presented most frequently (264/362, 72.9%), followed by a family member (197/324, 60.8%) and television (57/162, 35.2%). However, in Hefei, seeking doctors' or health specialists' advice remained the first option (112/151, 74.2%), followed by a family member (85/169, 50.3%) and friends and colleagues (69/167, 41.3%). Multimedia Appendix 3 displays the preferred media for accessing food risk and related information in Beijing and Hefei.

An average of 83.6% (n=2584) of Chinese respondents heard at least one food rumor, which included 95.7% (n=1462) from Beijing and 71.8% (n=1122) from Hefei. The respondents in Beijing were confronted more frequently by food rumors (range 346-1253) than those in Hefei (range 155-946). Interpersonal connection was the most frequent communication source (698/1253, 55.7%), followed by traditional media (325/1253, 25.9%) and internet portals (144/1253, 11.5%).

The K-means cluster identified 2 groups of users for learning narratives regarding food rumors. To be specific, in the high conspiracy belief group in Beijing, the category of interpersonal connection had the highest frequency of learning food rumor (130/275, 47.3%), followed by traditional media (72/275, 26.2%) and internet portals (49/275, 17.8%). A chi-square test of independence examined the conspiracy belief and media sources, and showed statistical significance (χ^2_5=20.08 [n=1253]; P=.001). A similar trend was also observed for the respondents in Hefei. For high conspiracy belief respondents, the category of interpersonal communication was the most frequent media source (182/346, 52.6%), followed by traditional media (113/346, 32.7%) and internet portals (31/346, 9.0%). However, there was no statistical significance for the high and low conspiracy beliefs in Hefei. Table 3 displays the crosstab comparison of conspiracy beliefs with access to the different categories of media sources in Beijing and Hefei in 2017.

Table 3. Crosstab comparison of conspiracy beliefs with access to different media sources in Beijing and Hefei.

Media source	Beijing, n (%)			Hefei, n (%)		
	Total (N=1253)	Low conspiracy belief (n=978)	High conspiracy belief (n=275)	Total (N=946)	Low conspiracy belief (n=600)	High conspiracy belief (n=346)
Interpersonal media	698 (55.7)	568 (58.1)	130 (47.3)	534 (56.4)	352 (58.7)	182 (52.6)
Public organizations	12 (1.0)	10 (1.0)	2 (0.7)	8 (0.8)	5 (0.8)	3 (0.9)
Traditional media	325 (25.9)	253 (25.9)	72 (26.2)	280 (29.6)	167 (27.8)	113 (32.7)
Internet portals	144 (11.5)	95 (9.7)	49 (17.8)	82 (8.7)	51 (8.5)	31 (9.0)
Social media	72 (5.7)	51 (5.2)	21 (7.6)	32 (3.4)	16 (2.7)	16 (4.6)
Others	2 (0.2)	1 (0.1)	1 (0.4)	10 (1.1)	9 (1.5)	1 (0.3)

The top publicized food rumor in Beijing was counterfeit seaweed (mean score 2.98, SD 1.05), followed by GMF involving chicken with 6 wings (mean score 2.69, SD 0.83) and seedless grapes with contraceptives (mean score 2.62, SD 0.96). In comparison, the most highly publicized rumor in Hefei was counterfeit seaweed (mean score 2.86, SD 0.98), followed by

seedless grapes with contraceptives (mean score 2.65, SD 0.86) and GMF involving chicken with 6 wings (mean score 2.62, SD 0.90). Table 4 outlines the analysis of food fraud with rumors and its correlation with consumer conspiracy beliefs in Beijing and Hefei.

Table 4. Correlation analysis of food fraud with rumors and consumer trust in Beijing and Hefei.

Food rumors and related terms	Beijing (n=1462), mean score (SD)	Hefei (n=1122), mean score (SD)	Correlation
1. Food counterfeit: seaweed	2.98 (1.05)	2.86 (0.98)	0.044
2. Addictive: strawberry	2.12 (0.82)	2.49 (0.79)	−0.012
3. Safety: microwaved food	2.45 (0.93)	2.60 (0.84)	−0.050
4. Implicated food: instant noodles	1.98 (0.84)	2.28 (0.84)	−0.103[a]
5. Contaminated: crayfish	2.45 (1.03)	2.47 (0.96)	−0.013
6. Safety: hookworm in pork	2.32 (0.97)	2.62 (0.99)	−0.076
7. GMF[b]: Six-wing chicken	2.69 (0.83)	2.62 (0.90)	−0.046
8. Authenticity: seedless grapes with contraceptives	2.62 (0.96)	2.65 (0.86)	−0.004

[a]$P<.01$.

[b]GMF: genetically modified food.

The K-means cluster identified 2 groups of distrust levels for learning narratives regarding food rumors. All of the narratives regarding food fraud with rumors showed a statistical significance between users with a low level of distrust and their counterparts with a high level of distrust in Beijing and Hefei.

Specifically, more users with a higher score indicated a high distrust of food rumors, such as "counterfeit seaweed" (202/638, 31.7%) and "implicated food of instant noodles" (275/1253, 21.9%) in Beijing. In comparison, a low distrust of food rumors, such as the implicated food narrative "instant noodles are junk food" (600/946, 63.4%) and another narrative "crayfish are genetically modified crops which deals with corpses and grows in unsanitary water with exceeded heavy metals" (266/296, 89.9%), was prevalent in Hefei. Interestingly, much more users in Hefei distrusted the rumor narrative "big and sweet strawberries are with excessive swelling agent" (235/547, 43.0%) than their counterparts in Beijing (27/800, 3.5%). Another rumor regarding the use of genetically modified chicken with 6 wings by KFC also had contrasting trust levels in the 2 cities, with lower distrust in Beijing (94/794, 11.8%) and higher distrust in Hefei (182/449, 40.5%). Table 5 presents the findings of univariate ANOVA comparing the food items in frauds and rumors perceived by users with different levels of distrust.

Table 5. Comparison and analysis of distrust levels of rumors among users in Beijing and Hefei.

Food fraud and related terms	Beijing users				Hefei users			
	Low, n (%)	High, n (%)	F test (df)	P value	Low, n (%)	High, n (%)	F test (df)	P value
1. Food counterfeit: seaweed	436 (68.3)	202 (31.7)	12277.62 (725)	<.001	253 (81.6)	57 (18.4)	259.67 (308)	<.001
2. Additive: strawberry	773 (96.5)	27 (3.5)	214.67 (801)	<.001	312 (57.0)	235 (43.0)	1828.17 (545)	<.001
3. Safety: microwaved food	646 (90.2)	70 (9.8)	533.26 (714)	<.001	279 (88.9)	35 (11.1)	216.77 (312)	<.001
4. Implicated food/safety: instant noodles	978 (78.1)	275 (21.9)	2392.72 (1251)	<.001	600 (63.4)	346 (36.6)	2244.68 (944)	<.001
5. Contamination/quality/genetically modified food: crayfish	381 (87.4)	55 (12.6)	374.62 (434)	<.001	266 (89.9)	30 (10.1)	207.05 (294)	<.001
6. Safety/authenticity: pork hookworm	426 (88.6)	55 (11.4)	383.50 (479)	<.001	216 (85.4)	37 (14.6)	20.51 (10)	<.001
7. KFC GMF[a]: Six-wing chicken	700 (88.2)	94 (11.8)	616.90 (792)	<.001	267 (59.5)	182 (40.5)	1047.60 (447)	<.001
8. Authenticity: grapes with contraceptives	302 (87.3)	44 (12.7)	274,60 (344)	<.001	138 (89.0)	17 (11.0)	112.21 (153)	<.001

[a]GMF: genetically modified food.

The terminology within the domain of food fraud and rumors included the following: food authenticity, food additive, food safety, food contamination, implicated food, and questionable GMF sources. The most significant misinformation of food fraud in rumors included "seaweed is made out of plastic" ($\chi^2_4=29.26$; $P<.001$), "instant noodles are junk food" ($\chi^2_4=20.08$; $P<.001$), "crayfish are GM crops" ($\chi^2_4=29.87$; $P<.001$), "hookworm in pork" ($\chi^2_4=40.18$; $P<.001$), and "seedless grapes were sprayed with contraceptives" ($\chi^2_4=19.95$; $P<.001$) in Beijing. In comparison, only 1 issue regarding the rumor "seaweed is made of plastic" with different levels of distrust attitudes showed statistical significance in Hefei ($\chi^2_4=17.66$;

P=.003). Multimedia Appendix 4 shows the comparative analysis of the groups (low and high distrust) for learning about food rumors by accessing different media sources in Beijing and Hefei.

The top publicized food incidents perceived in Beijing were branded contaminated yogurt from Vietnam (mean score 4.08, SD 1.27), followed by suspected tainted tap water (mean score 4.02, SD 1.29) and expired milk powder from New Zealand (mean score 3.99, SD 1.31). In comparison, the most highly publicized news exposure in Hefei was the use of poppy shells in a chain restaurant in Beijing (mean score 4.12, SD 1.09), followed by a tainted sesame product (mean score 4.02, SD 1.12) and branded contaminated yogurt from Vietnam (mean score 4.01, SD 1.16). Table 6 outlines the analysis of food fraud items in scandals correlated with measuring consumer conspiracy beliefs between Beijing and Hefei.

Table 6. Food items and incident types involving consumer trust in Beijing and Hefei.

Food incident and related terms	Beijing (n=1462), mean score (SD)	Hefei (n=1122), mean score (SD)	Correlation
1. Implicated foods/abnormal conditions: fish out of shelves	3.67 (1.34)	4.03 (1.04)	−0.033
2. Food quality/security: delivery	3.19 (1.46)	3.53 (1.22)	−0.073[a]
3. Food adulterant: beef with duck meat	3.73 (1.37)	3.96 (1.13)	−0.060[a]
4. Food crime: poppy shells	3.74 (1.40)	4.12 (1.09)	−0.080[b]
5. Food authenticity: fake milk powder	3.86 (1.34)	3.49 (1.29)	−0.040
6. Food contaminant/security: radiation in seafood	3.65 (1.39)	3.47 (1.31)	−0.040
7. Food quality: expired dairy products	3.99 (1.31)	3.95 (1.13)	−0.049
8. Food contaminant: tainted sesame	3.86 (1.35)	4.02 (1.12)	−0.046
9. Food quality: frozen meat	3.90 (1.37)	3.81 (1.24)	−0.068[a]
10. Food quality/contaminant: yogurt	4.08 (1.27)	4.01 (1.16)	−0.019
11. Food contaminant: suspected tainted water	4.02 (1.29)	3.95 (1.16)	−0.022
Subtotal	3.79 (1.00)	3.85 (0.80)	−0.082[b]

[a]*P*<.05.

[b]*P*<.01.

The food fraud incidents and related terminology included food authenticity, food additive, food adulterant, food contamination, implicated food, food quality, and concerned sources with GMFs. The K-means cluster also identified 2 groups with distrust levels (low vs high) against negative food reports. All of the narratives regarding food fraud with scandals showed a significant difference between low-level distrust users and their high-level distrust counterparts in Beijing and Hefei.

Overall, responses to the food fraud and incident risks were lower in Beijing than in Hefei (*r*=−0.082; *P*=.006). The following 4 scandals reflected different perceptions between dwellers in Beijing and Hefei: food hygiene on an online food ordering platform (*r*=−0.073; *P*=.014), food adulteration by mixing beef with duck meat (*r*=−0.060; *P*=.046), food crime by using poppy shells (*r*=−0.080; *P*=.007), and food quality in problematic frozen meat (*r*=−0.068; *P*=.02). Multimedia Appendix 5 shows the comparative analysis of the groups (low and high distrust) for accessing media sources in learning about food in fraud and negative reports in Beijing and Hefei.

Discussion

Principal Findings

The main findings of this study were that around 73.6% (out of 2584) of Chinese respondents preferred to go to their physicians for quarrying food health information first; however, when asked where they actually went and got access to food rumors, up to 36.6% (out of 1462) of Beijing respondents and 55.6% (out of 1122) of Hefei respondents reported going online first. Interpersonal connection had the highest frequency among communication sources (698/1253, 55.7%). There was a significant relationship between media use and media category in Beijing (*P*<.001) and Hefei (*P*<.001). Overall, responses to the food fraud and incident risks were lower in Beijing than in Hefei (*P*=.006). The urban dwellers in Beijing and their rural counterparts in Hefei also differed in terms of perceiving different levels of food risk from different media sources.

Foods or raw ingredients most likely to be targeted for adulteration or fraud are those of high economical value to the diet, which are subject to the vagaries of weather during their growth, harvest, storage, or transport. Food quality and food safety incidents involving criminal conduct are considered as intentional acts, and there is zero tolerance from the food regulators in China. Consumers have to be on guard against not only fake food, but also spreading fake information and rumors about food.

Most Chinese respondents were principally confronted with food rumors through interpersonal connections, followed by traditional media and internet portals. Believing in conspiracies supports alternative views to construct radical beliefs in social uncertainty, and laypeople are now creating new articulations of discourse in the public sphere. This is because of their

innovative and often subversive language. Despite the characterization of conspiracy beliefs as paranoid by some in public discourse, they remain robust sources of skepticism regarding important public health recommendations able to prevent the spread of rumors.

For example, a food rumor regarding seafood products being exposed to radiation has been highly publicized. Many countries, including China, have banned seafood imports from Fukushima and other Japanese prefectures due to contamination of surrounding waters [45]. Further, the massive 2011 Fukushima nuclear disaster led to dramatic price drops in Japanese seafood. Consumers were warned that radiation continues to negatively impact marine life in the vicinity of Fukushima. Thus, many consumers believe that seafood products have high levels of radiation that can cause irreversible damage to human cells. Although 10 years have passed since the disaster, the appetite of many consumers of Japanese seafood has remained low.

In our analysis of the prevalence of food incident–related conspiracy beliefs, we found that acceptance of those beliefs highlighted the narratives of seafood contaminants for Beijing respondents (mean score 3.65, SD 1.39) and the narratives of food crime in adding poppy shells for Hefei respondents (mean score 4.12, SD 1.09). Overall, the level of conspiracy belief acceptance grew over the issues of food quality and contamination in Beijing, while the level of acceptance was increasing over the issues of food crime and contamination in Hefei. As with many conspiracies, these beliefs can be accepted by the same person despite their logical incompatibility. It is argued that the underlying distrust of food supply and chain systems in a local market rather than merely the consistency of the media content appears to motivate their acceptance.

It may be difficult to understand how people could believe the rumor that seaweed is made out of plastic or chickens used by the KFC fast food chain were genetically modified to have 6 wings and 8 legs. Nonetheless, food fraud involving some food category or with specific global brands (eg, KFC, Abbott, and Beingmate) has become an issue influencing conspiracy belief and causing economic loss. The rumors appeared in posts on WeChat, and the defamatory messages were read widely. The rumors are just some of many fake food stories going viral, where fears about the safety of food products have become deep-seated in the wake of major cases of food contamination, and food framed as artificial and dangerous may also function as a counterpoint to promote safe and sustainable food in China [46].

Weibo and WeChat were found to be the most used social media for accessing food information in China. For example, a video clip of news was circulated widely and had morphed into dozens of versions, accruing more than 2 million views on Weibo, China's most popular social media platform. Nonetheless, there is an abundance of social media in China without gatekeepers for polarized communities within which antagonistic spheres are created, although they do foster engagement with food fraud discussion as well.

Limitations

Although the HINTS-China provides evidence of media use in the persistence of conspiracy beliefs regarding food fraud rumors, it is almost impossible to disentangle the actual risks perceived from media exposure by researchers. Concomitantly, because we rely on self-reports of recall behavior, we cannot confirm that those reports reflect actual behavior. We also lack evidence to claim purposeful information-seeking behavior and the scanning use of health risk information on food fraud, since our study solely relied on a self-report method.

Another limitation that needs to be recognized is that it is not feasible to generalize the national probability sample to the Chinese population since the survey is only applicable to changes that occurred in 2017. The types of food fraud incidents in China are underrepresented. The effects of media use on food fraud conspiracy beliefs beyond that period remain to be studied. Finally, despite the ability to observe differences in conspiracy beliefs associated with media use in 2 cities, we cannot make strong causal claims because it still remains possible that characteristics other than those for which we controlled drove the changes in those beliefs.

Conclusion

Passive media information seekers or those with information-scanning behavior were the majority who were exposed to information that was gathered incidentally from sources within their environment. Consequently, majority of respondents in Beijing and Hefei preferred to use interpersonal connection for learning food safety concerns, while very few respondents reported access to media run by public organizations for the same purpose. Nonetheless, the food fraud narratives examined according to the conspiracy belief showed that social media play important roles in influencing the attitude against negative reports, and 9 out of 11 food incidents tended to be perceived with higher conspiracy belief among Hefei respondents.

Food fraud is characterized with the intent to harm and is mainly done for economic gain. Thus, the main events of food fraud that occurred in 2016, specifically those associated with food rumors and incidents, were summarized in this study. The typical food fraud covered food adulteration, authenticity, contamination, crime, integrity, protection, quality, and safety, which were used as prompts for attitudinal and perceptual elicitation. Many food rumors place particular emphasis on compositional aspects of food, such as texture, color, and shape. They also involve food origin, geographical consideration, rearing or production systems, processing, and storage, among others. Food fraud regarding food originating from Japan, Vietnam, and New Zealand implies that the information concerning food fraud in these countries may negatively affect the valuation of import locations.

The food rumor narratives that examined the conspiracy belief of distrust degree showed that consumers with prior knowledge of food fraud incidents decreased the valuation of the sources less when they received further information about food fraud. Fueled by constant access to mobile devices, daily online media consumption has increased steadily since 2017. Chinese adults

tend to spend more time each day listening to, reading, watching, or interacting with online media; however, their voracious appetite for digital content exists alongside a continued fondness for traditional media outlets. Furthermore, the results implied that prior consumer knowledge and later response to fraudulent behavior in a product can spread to other products.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of food fraud narratives and health risks of food incidents.
[DOCX File , 27 KB - formative_v6i3e32302_app1.docx]

Multimedia Appendix 2
Media sources.
[DOCX File , 14 KB - formative_v6i3e32302_app2.docx]

Multimedia Appendix 3
Sources for accessing food risk information.
[DOCX File , 19 KB - formative_v6i3e32302_app3.docx]

Multimedia Appendix 4
Comparative analysis of the groups (low and high distrust) for learning about food rumors by accessing different media sources in Beijing and Hefei.
[DOCX File , 26 KB - formative_v6i3e32302_app4.docx]

Multimedia Appendix 5
Comparative analysis of the groups (low and high distrust) for accessing media sources in learning about food in fraud and negative reports in Beijing and Hefei.
[DOCX File , 31 KB - formative_v6i3e32302_app5.docx]

References

1. Spink J, Moyer DC. Defining the public health threat of food fraud. J Food Sci 2011;76(9):R157-R163. [doi: 10.1111/j.1750-3841.2011.02417.x] [Medline: 22416717]
2. Kendall H, Kuznesof S, Dean M, Chan M, Clark B, Home R, et al. Chinese consumer's attitudes, perceptions and behavioural responses towards food fraud. Food Control 2019 Jan;95:339-351. [doi: 10.1016/j.foodcont.2018.08.006]
3. Spink J, Embarek PB, Savelli CJ, Bradshaw A. Global perspectives on food fraud: results from a WHO survey of members of the International Food Safety Authorities Network (INFOSAN). NPJ Sci Food 2019 Jul 17;3(1):12 [FREE Full text] [doi: 10.1038/s41538-019-0044-x] [Medline: 31341950]
4. Zhang W, Xue J. Economically motivated food fraud and adulteration in China: An analysis based on 1553 media reports. Food Control 2016 Sep;67:192-198. [doi: 10.1016/j.foodcont.2016.03.004]
5. Spink J, Ortega DL, Chen C, Wu F. Food fraud prevention shifts the food risk focus to vulnerability. Trends in Food Science & Technology 2017 Apr;62:215-220. [doi: 10.1016/j.tifs.2017.02.012]
6. Pei X, Tandon A, Alldrick A, Giorgi L, Huang W, Yang R. The China melamine milk scandal and its implications for food safety regulation. Food Policy 2011 Jun;36(3):412-420. [doi: 10.1016/j.foodpol.2011.03.008]
7. Lami F, Radhi F, Al Dahhan S, Hashim RA, Mahmood H, Araj R, et al. Knowledge, Attitude, and Practices of Food Handlers on Food Safety and Personal Hygiene During Arbaeenia Mass Gathering, Baghdad, Iraq, 2014: Cross-Sectional Study. JMIR Public Health Surveill 2019 Oct 09;5(4):e10922 [FREE Full text] [doi: 10.2196/10922] [Medline: 31599735]
8. Liu P, Ma L. Food scandals, media exposure, and citizens' safety concerns: A multilevel analysis across Chinese cities. Food Policy 2016 Aug;63:102-111. [doi: 10.1016/j.foodpol.2016.07.005]
9. Mian A, Khan S. Coronavirus: the spread of misinformation. BMC Med 2020 Mar 18;18(1):89 [FREE Full text] [doi: 10.1186/s12916-020-01556-3] [Medline: 32188445]
10. Douglas KM, Uscinski JE, Sutton RM, Cichocka A, Nefes T, Ang CS, et al. Understanding Conspiracy Theories. Political Psychology 2019 Mar 20;40(S1):3-35. [doi: 10.1111/pops.12568]
11. HINTS-China: Promoting Understanding about Consumers' Health Information Needs and Practices in China. Health Information National Trends Survey. URL: https://hints.cancer.gov/docs/HINTS-Kreps_508c.pdf [accessed 2020-12-28]
12. Romer D, Jamieson KH. Conspiracy theories as barriers to controlling the spread of COVID-19 in the U.S. Soc Sci Med 2020 Oct;263:113356 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113356] [Medline: 32967786]

XSL•FO
RenderX

13. Allcott H, Gentzkow M, Yu C. Trends in the diffusion of misinformation on social media. Research & Politics 2019 May 09;6(2):205316801984855. [doi: 10.1177/2053168019848554]

14. Bursztyn L, Rao A, Roth C, Yanagizawa-Drott D. Misinformation During a Pandemic. NBER. URL: https://www.nber.org/system/files/working_papers/w27417/w27417.pdf [accessed 2022-02-06]

15. Swire-Thompson B, Lazer D. Public Health and Online Misinformation: Challenges and Recommendations. Annu Rev Public Health 2020 Apr 02;41(1):433-451. [doi: 10.1146/annurev-publhealth-040119-094127] [Medline: 31874069]

16. Oleksy T, Wnuk A, Maison D, Łyś A. Content matters. Different predictors and social consequences of general and government-related conspiracy theories on COVID-19. Pers Individ Dif 2021 Jan 01;168:110289 [FREE Full text] [doi: 10.1016/j.paid.2020.110289] [Medline: 32834288]

17. van der Wal RC, Sutton RM, Lange J, Braga JPN. Suspicious binds: Conspiracy thinking and tenuous perceptions of causal connections between co-occurring and spuriously correlated events. Eur J Soc Psychol 2018 Dec 26;48(7):970-989 [FREE Full text] [doi: 10.1002/ejsp.2507] [Medline: 30555189]

18. Uscinski JE, Klofstad C, Atkinson MD. What Drives Conspiratorial Beliefs? The Role of Informational Cues and Predispositions. Political Research Quarterly 2016 Jan 13;69(1):57-71. [doi: 10.1177/1065912915621621]

19. Bricker BJ. Climategate: A Case Study in the Intersection of Facticity and Conspiracy Theory. Communication Studies 2013 Apr;64(2):218-239. [doi: 10.1080/10510974.2012.749294]

20. Goertzel T. Conspiracy theories in science. EMBO Rep 2010 Jul 11;11(7):493-499 [FREE Full text] [doi: 10.1038/embor.2010.84] [Medline: 20539311]

21. Lahrach Y, Furnham A. Are modern health worries associated with medical conspiracy theories? J Psychosom Res 2017 Aug;99:89-94. [doi: 10.1016/j.jpsychores.2017.06.004] [Medline: 28712435]

22. Guo Q, Yao N, Zhu W. How consumers' perception and information processing affect their acceptance of genetically modified foods in China: A risk communication perspective. Food Res Int 2020 Nov;137:109518. [doi: 10.1016/j.foodres.2020.109518] [Medline: 33233152]

23. Lee GE. The Establishment of Ontario's Local Health Integration Networks: A Conflation of Regionalization with Integration of Health Services. HRO-ORS 2018 Nov 02;6(2):2850. [doi: 10.13162/hro-ors.v6i2.2850]

24. Liang J, Liu X, Zhang W. Public Underst Sci 2019 Nov 24;28(8):991-1004. [doi: 10.1177/0963662519876782] [Medline: 31547775]

25. Wang X, Song Y. Viral misinformation and echo chambers: the diffusion of rumors about genetically modified organisms on social media. INTR 2020 Jun 22;30(5):1547-1564. [doi: 10.1108/intr-11-2019-0491]

26. Zucker HA. Tackling Online Misinformation: A Critical Component of Effective Public Health Response in the 21st Century. Am J Public Health 2020 Oct;110(S3):S269-S269. [doi: 10.2105/ajph.2020.305942]

27. Spink J, Bedard B, Keogh J, Moyer DC, Scimeca J, Vasan A. International Survey of Food Fraud and Related Terminology: Preliminary Results and Discussion. J Food Sci 2019 Oct 23;84(10):2705-2718. [doi: 10.1111/1750-3841.14705] [Medline: 31546281]

28. Beckjord EB, Finney Rutten LJ, Squiers L, Arora NK, Volckmann L, Moser RP, et al. Use of the internet to communicate with health care providers in the United States: estimates from the 2003 and 2005 Health Information National Trends Surveys (HINTS). J Med Internet Res 2007 Jul 12;9(3):e20 [FREE Full text] [doi: 10.2196/jmir.9.3.e20] [Medline: 17627929]

29. Nelson DE, Kreps GL, Hesse BW, Croyle RT, Willis G, Arora NK, et al. The Health Information National Trends Survey (HINTS): development, design, and dissemination. J Health Commun 2004 Sep;9(5):443-60; discussion 81. [doi: 10.1080/10810730490504233] [Medline: 15513791]

30. Chou WS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009 Nov 27;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]

31. Arora A, Bansal S, Kandpal C, Aswani R, Dwivedi Y. Measuring social media influencer index- insights from facebook, Twitter and Instagram. Journal of Retailing and Consumer Services 2019 Jul;49:86-101. [doi: 10.1016/j.jretconser.2019.03.012]

32. Keles B, McCrae N, Grealish A. A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. International Journal of Adolescence and Youth 2019 Mar 21;25(1):79-93. [doi: 10.1080/02673843.2019.1590851]

33. Ruparel N, Dhir A, Tandon A, Kaur P, Islam JU. The influence of online professional social media in human resource management: A systematic literature review. Technology in Society 2020 Nov;63:101335. [doi: 10.1016/j.techsoc.2020.101335]

34. Bui Thi Thu H, Pham Van T, Mirzoev T. Implementation of Health Policy on Establishment of Provincial Center of Diseases Control (CDC) in Vietnam. RMHP 2020 Jul;Volume 13:915-926. [doi: 10.2147/rmhp.s250748]

35. Wang Y, McKee M, Torbica A, Stuckler D. Systematic Literature Review on the Spread of Health-related Misinformation on Social Media. Soc Sci Med 2019 Nov;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]

36. Zhang L, Jiang S. Linking health information seeking to patient-centered communication and healthy lifestyles: an exploratory study in China. Health Educ Res 2021 Apr 12;36(2):248-260. [doi: 10.1093/her/cyab005] [Medline: 33544831]

37.   Quattrociocchi W. Inside the Echo Chamber. Sci Am 2017 Mar 14;316(4):60-63. [doi: 10.1038/scientificamerican0417-60] [Medline: 28296846]

38.   Han G, Liu Y. Does Information Pattern Affect Risk Perception of Food Safety? A National Survey in China. Int J Environ Res Public Health 2018 Sep 05;15(9):1935 [FREE Full text] [doi: 10.3390/ijerph15091935] [Medline: 30189673]

39.   Kreps GL, Rutten LF. Building the evidence base in cancer communication: next steps. In: Rutten IF, Hesse B, Moser R, Kreps GL, editors. Building the evidence base in cancer communication. Cresskill, NJ: Hampton Press; 2011:315-322.

40.   Kreps GL, Yu G, Zhao X, Chou SW, Xu Z, Song M. Extending the US Health Information National Trends Survey to China and beyond: promoting global access to consumer health information needs and practices. In: Kreps GL, Dini P, editors. Global health 2012: the first international conference on global health challenges. Wilmington, DE: International Academy, Research, and Industry Association; 2012:119-122.

41.   Kreps GL, Yu G, Zhao X, Chou SW, Hesse B. Expanding the NCI Health Information National Trends Survey From the United States to China and Beyond: Examining the Influences of Consumer Health Information Needs and Practices on Local and Global Health. Journalism & Mass Communication Quarterly 2017 Jan 13;94(2):515-525. [doi: 10.1177/1077699016687725]

42.   Zhao X, Mao Q, Kreps GL, Yu G, Li Y, Chou SW, et al. Cancer information seekers in china: a preliminary profile. J Health Commun 2015 Apr 06;20(5):616-626. [doi: 10.1080/10810730.2015.1012244] [Medline: 25845031]

43.   Zhang L, Qin Y, Li P. Media Complementarity and Health Information Acquisition: A Cross-sectional Analysis of the 2017 HINTS-China Survey. J Health Commun 2020 Apr 02;25(4):291-300. [doi: 10.1080/10810730.2020.1746868] [Medline: 32255740]

44.   People's Daily. China Science Communication. URL: http://kpzg.people.com.cn/ [accessed 2020-12-28]

45.   Import Alert 99-33. US Food & Drug Administration. URL: https://www.accessdata.fda.gov/cms_ia/importalert_621.html [accessed 2020-12-21]

46.   Vargas Meza X, Yamanaka T. Food Communication and its Related Sentiment in Local and Organic Food Videos on YouTube. J Med Internet Res 2020 Aug 10;22(8):e16761 [FREE Full text] [doi: 10.2196/16761] [Medline: 32773370]

Abbreviations

ANOVA: analysis of variance
CFDA: China Food & Drug Administration
GMF: genetically modified food
HINTS: Health Information National Trends Survey
HINTS-China: Health Information National Trends Survey in China

XSL•FO
RenderX

<u>Original Paper</u>

Patients' Perspectives About the Treatment They Receive for Cardiovascular Diseases and Mental Disorders: Web-Based Survey Study

Philippe Courtet[1], MD; Catherine Pecout[2], MSc; Anne-Félice Lainé-Pellet[2], DPharm; Michael Chekroun[3], BSc, MBA; Charlotte Avril[3], MSc, DPharm; Jean-Jacques Mourad[4], MD

[1]Service d'Urgence et Post-Urgence Psychiatrique, Centre Hospitalier Universitaire de Montpellier, Montpellier, France

[2]Viatris, Pfizer Inc France, Paris, France

[3]Carenity, ELSE CARE, Paris, France

[4]Département de Médecine Interne et Centre d'Excellence, Société Européenne d'Hypertension, Groupe Hospitalier Paris Saint Joseph, Paris, France

Corresponding Author:
Philippe Courtet, MD
Service d'Urgence et Post-Urgence Psychiatrique
Centre Hospitalier Universitaire de Montpellier
371 avenue du Doyen Gaston Giraud
Montpellier, 34295
France
Phone: 33 0467336733
Email: philippe.courtet@univ-montp1.fr

Abstract

Background: Noncommunicable disease (NCD)–related deaths account for 71% of deaths worldwide. The World Health Organization recently developed a global action plan to address the impact of NCDs, with the goal of reducing the number of premature NCD-related deaths to 25% by the year 2025. Appropriate therapeutic adherence is critical for effective disease management; however, approximately 30%-50% of patients with an NCD do not comply with disease management activities as prescribed. Web-based patient communities can represent platforms from which specific information on patients' perception of treatment adherence can be gathered outside of a clinical trial setting.

Objective: This study aims to better understand patients' perspectives regarding therapeutic adherence and iatrogenic risk in 2 major groups of NCDs for which poor disease management can have fatal consequences: cardiovascular diseases and mental disorders. Therapeutic adherence, motivational factors, patients' awareness and perception of iatrogenesis, and treatment tools used by patients were assessed.

Methods: A web-based survey was performed among patients with cardiovascular diseases or mental disorders or both conditions who were registered on the French Carenity platform, a web-based community in which patients with an NCD can share experiences and receive support and information. The study inclusion criteria were defined as follows: diagnosis of cardiovascular disease or mental disorder or both conditions (self-declared), age ≥18 years, residence in France, registration on the French Carenity platform, and ongoing pharmaceutical treatment for the condition. Patients who met the inclusion criteria were then invited to complete a self-administered web-based questionnaire that included questions addressing therapeutic adherence and iatrogenic risk.

Results: A total of 820 patients were enrolled in the study, including patients with cardiovascular diseases (403/820, 49.2%), patients with mental disorders (292/820, 35.6%), and patients with both cardiovascular diseases and mental disorders (125/820, 15.2%). The mean age of the participants was 55.2 (SD 12.7) years. We found that 82.8% (679/820) of patients experienced adverse effects of medication. Patients tended to perceive themselves to be more adherent than they actually were; a significant number of patients disregarded their prescription and stopped or interrupted medication without consulting with a doctor. Patients with cardiovascular diseases were nearly twice as adherent as patients with a mental disorder ($P \le .001$). Adherence was significantly associated with gender ($P \le .001$), age ($P \le .001$), and treatment complexity ($P \le .001$). Finally, for each disease type, 3 patient profiles were identified, which provide interesting insight for improving therapeutic adherence and adjustment strategies specifically according to patient behavior.

Conclusions: This study provides insight into the perspectives of patients receiving therapy for cardiovascular diseases or mental disorders or both conditions, which could help improve the management of NCDs and prevent premature death. Our study also

shows that web-based patient platforms provide new opportunities to improve disease management by understanding patients' experiences.

Introduction

Deaths from noncommunicable diseases (NCDs), also known as chronic diseases, account for approximately 41 million deaths each year and 71% of deaths worldwide [1]. The World Health Organization recently developed a global action plan to address the impact of NCDs, with a goal of reducing premature NCD-related deaths to 25% by 2025 [2]. As premature deaths due to NCDs are avoidable, many policy recommendations have been proposed to prevent NCDs and improve disease management for patients with NCDs.

Cardiovascular disease–related deaths account for the majority of NCD-related deaths and 30% of global mortality [3]. Mental disorders also represent a major group of NCDs and account for approximately 14% of the global disease burden [4]. Cardiovascular diseases and mental disorders are major economic burdens on health care systems in terms of the direct (eg, medical consultations, hospitalizations, rehabilitation services, and medications) and indirect (eg, loss of productivity and short- or long-term disability) costs associated with mortality and morbidity [4-6].

Appropriate therapy management, including medication adherence, is critical for effective disease management and for improving patients' overall quality of life [7,8]. However, approximately 30%-50% of patients with an NCD do not comply with disease prevention and management activities [9,10] such as following treatment as prescribed by a physician, staying up-to-date with medical appointments, engaging in regular physical activity, and making necessary dietary changes [7,11]. Medication adherence is defined as the extent to which patients take medications as prescribed in agreement with their health care provider [12]. Poor therapeutic adherence, which has significant effects on treatment outcome and disease prognosis [13], is driven by many factors such as limited disease awareness, poor understanding of the benefits and efficacy of prescribed regimens, and perceived or actual barriers (eg, adverse effects, financial constraints) [7,9,13,14]. Poor disease management can have fatal consequences, especially in patients with cardiovascular diseases and mental disorders [2,15,16]. Furthermore, patients with an NCD are at risk of iatrogenic disease, which is any pathologic condition caused by adverse medication reactions or complication induced by nondrug medical interventions, including diagnosis, intervention, error, and negligence. Although iatrogenic diseases can have major psychomotor and social consequences, most of these are avoidable with close disease management [17,18].

Few studies have been performed on therapeutic adherence from the patient point of view, which could provide valuable insight into patients' perspectives and experiences for improving NCD management. As patients involved in adherence studies are often not representative of the general patient population for a given disease, bias may limit result extrapolation. However, web-based platforms on which patients can obtain information and share their medical experiences anonymously, without medical supervision, may more accurately represent patients' perspectives and behaviors.

Web-based patient platforms such as registries, forums, social networks, and web-based communities offer patients the opportunity to participate in scientific studies and voluntarily share experiences regarding treatment benefits and burdens outside of the clinical setting [19]. Such information provides researchers with a better understanding of patients' experiences, expectations, and unmet needs [19]. The insight gained from these web-based resources may also enhance clinical decision making, study protocol development, and patient recruitment [19]. The Carenity platform is an international web-based community for patients with chronic diseases and their caregivers. The platform provides an environment for patients to share their experiences, monitor their health, provide support, and contribute to medical research through web-based surveys [20]. Currently, over 400,000 patients (88%), primarily with chronic diseases, and their caregivers (12%) from 6 countries (France, Italy, Germany, Spain, the United Kingdom, and the United States) are registered on the platform.

Most studies on therapeutic adherence have been based on in-person interviews. However, under these conditions, patients may be reluctant to report poor medication adherence to avoid disappointing their physician. As social media networks represent useful and valuable resources for patients to obtain medical information and share their experiences with other patients, a web-based survey of patients with cardiovascular diseases or mental disorders or both conditions was performed to better understand patients' perspectives of therapeutic adherence and iatrogenic risk. An anonymous and self-administered questionnaire was used to understand aspects of therapeutic adherence, including motivational factors, patients' awareness and perception of iatrogenesis, and tools used by patients to facilitate adherence.

Methods

Study Design

A web-based survey of patients with cardiovascular diseases or mental disorders or both conditions was conducted. All patients

were registered on the French Carenity platform, a web-based community in which both patients with chronic diseases and their caregivers can share their experiences, provide support, and share or receive information. A caregiver is defined as a person who provides care to someone with a chronic disease, disability, or other long-term health condition, typically outside a professional or formal framework. The community enhances patient-centered approaches by sharing patients' experiences through web-based surveys, in which members may voluntarily participate. In February 2020, approximately 7600 members with at least one mental disorder and 10,000 members with at least one cardiovascular disease were registered on the platform.

Participant Recruitment

Patients were recruited from February 14, 2020, to May 15, 2020, via invitation and follow-up no-reply emails. The study inclusion criteria were defined as follows: diagnosis of cardiovascular disease or mental disorder or both conditions (self-declared), age ≥18 years, residence in France, registration on the French Carenity platform, and ongoing pharmaceutical treatment for the condition. Patients who met the inclusion criteria and agreed to receive invitations were then contacted to complete a self-administered web-based questionnaire available on the Carenity website and promoted on Facebook. Patients who completed the questionnaire but did not meet the inclusion criteria were screened out, and patients who did not finish the questionnaire were not included in the analysis.

Data Collection

The questionnaire comprised 39 questions regarding sociodemographic and medical information, including questions addressing therapeutic adherence and iatrogenesis risk. The questionnaire was developed and approved by a multidisciplinary board of experts, including a psychiatrist and a cardiovascular specialist. It was also approved by 2 Carenity members to ensure that the proposed questions were appropriate for the target audience. Data collected on the Carenity platform are hosted in France on a secured computer server in accordance with the Commission Nationale de l'Informatique et des Libertés, declaration number no 1484083, dated March 29, 2011.

Demographic and Clinical Characteristics

Information on demographic variables (eg, age, gender, education level, and professional status) and clinical characteristics (eg, disease type, age at diagnosis, comorbidities, and current treatments) was collected.

Treatment Complexity Score

The treatment complexity score was calculated by assigning 1 point for each constraining aspect of the pharmacological therapy, such as "varying number of medications each day," "taking some medications with meals," "taking some medications outside mealtimes," "taking some medications at a set time each day," "taking some medications on certain days and not others," and "doses frequently change for some medications." The treatment complexity score ranged from 0 (simplest treatment) to 7 (most complex treatment), and complexity was then classified into 3 categories based on the score: simple (score 0-1), intermediate (score 2), and complex (score 3-7).

Patient Lifestyle and Risk Level

Information was collected on the following 4 behavioral risk factors most associated with NCDs: tobacco use, excessive alcohol consumption, unhealthy diet, and physical inactivity [2]. Height and weight data were also recorded for each patient to calculate BMI. Nutritional status was classified based on patient BMI (weight [kg]/height squared [m^2]) according to World Health Organization recommendations: underweight (BMI < 18.5), normal weight (BMI 18.5-24.9), overweight (BMI 25-29.9), and obese (BMI ≥ 30) [21].

Patients were then categorized into 3 risk level groups. The risk level of each patient was based on the following 5 risk factors: age ≥65 years, BMI ≥ 30, occasional or regular smoking, daily alcohol consumption or consumption of at least 6 drinks every week, and physical inactivity. The risk level was estimated by assigning 1 point per risk factor as follows: 0-1 risk factor, low risk; 2 risk factors, moderate risk; and 3-5 risk factors, high risk.

Iatrogenesis

To collect data on iatrogenic risk, questions focused on the adverse effects that patients experienced and the worries and fears that they had about their medication. Information was also collected on medications, including any interactions, adverse effects, benefits, and channels used to obtain information.

Medication Adherence

Perceived medication adherence was determined based on the response to the following question: "How well do you think you take your medications?" Patients were prompted to respond on a scale of 0 to 100 (0 corresponded to "I do not take my medications as prescribed" and 100 corresponded to "I take all my medications exactly as prescribed").

Actual adherence was assessed based on the response to the questions considering all medications and not only to those for cardiovascular diseases or mental disorders. Participants were asked the following 4 questions: "Regarding your medications, do you ever (1) take medication late or early, (2) disregard the prescribed dose without medical advice, (3) intentionally stop/interrupt a treatment without medical advice, and (4) unintentionally stop/interrupt a treatment (because you forgot, etc.)?" Respondents were instructed to select the most appropriate response (never, very rarely, sometimes, often), with only 1 response per question. The adherence score was then calculated by assigning points according to the frequency of nonadherent behaviors. For the question about taking medication late or early, points were assigned as follows: 1 point for very rarely, 2 points for sometimes, and 4 points for often. For the questions about disregarding the prescribed dose and unintentionally stopping or interrupting treatment, points were assigned as follows: 2 points for very rarely, 4 points for sometimes, and 6 points for often. Finally, for the question on intentionally stopping or interrupting treatment, points were assigned as follows: 4 points for very rarely, 8 points for sometimes, and 12 points for often. Patients were then grouped into the following 4 compliance categories based on the total points: perfectly adherent (0 points), mostly adherent (1-2 points), partially adherent (3-7 points), and poorly adherent (≥8 points).

Information was also collected on adherence reporting, including frequency of doctor notifications and reasons patients did not consult with their doctors. Finally, data were also collected on solutions implemented by patients to improve disease management, such as tools used and types of assistance received, and additional patient needs.

Statistical Analysis

Descriptive multivariable statistical analyses were performed. Categorical variables are expressed as absolute frequency and percentage. Continuous variable data are presented as the mean (SD) for normal distribution and as the median and interquartile range for non-normal distribution. A chi-square (χ^2) test was applied to determine statistically significant differences for categorical variables, and a one-way analysis of variance or t test was applied to determine statistically significant differences for continuous variables.

A multiple correspondence analysis was performed for each of the 2 disease types to identify profiles of patients with similar therapeutic behaviors. To refine these profiles, 3 unsupervised classification models (ascending hierarchical classification, Kmeans, and partitioning around medoids algorithm) were compared using indicators of similarity and dissimilarity. The method with the best results was the 3-class ascending hierarchical classification model based on 12 variables covering the following 3 categories: demographic variables (gender and age), medical characteristics (eg, number of medications, adherence, adverse effects, treatment adjustment, and complementary therapies), and survey responses (fear of disappointing the physician, fear of risks associated with the therapy, discussions with the pharmacist, awareness of the risks of adverse effects, and risk of experiencing adverse effects). Data processing and analysis were performed using R (version 3.6.1; R Core Team).

Patients with cardiovascular diseases only were classified into 3 different profiles: at-risk, reporter, and tolerant. Patients with mental disorders only were also grouped into 3 different profiles: fearful, at-risk, and confident.

Ethical Considerations

This survey was conducted in accordance with the Declaration of Helsinki and the principles of Good Clinical Practice. Prior to data collection, all patients provided consent on the Carenity website for the analysis of their anonymous health data for the study and subsequent publication of the findings (patients were informed that their health data would be collected and analyzed uniquely upon explicit consent, which was formalized by clicking the "Start" button at the bottom of the information page of the web-based survey). Participant privacy and confidentiality were guaranteed according to European laws and regulations (General Data Protection Regulation). As informed consent was provided by all patients prior to completing the survey, ethical review and approval were waived for this study.

Results

Respondent Profiles

During the recruitment period, 820 patients with mental disorders or cardiovascular diseases or both conditions were enrolled in the study. Depending on the disease category, the 820 patients were recruited, selected, and grouped according to the risk level, therapeutic adherence, and behavior profiles as shown in Figure 1. Patients who answered being a "patient with cardiovascular disease or a cardiovascular risk factor" in the questionnaire were categorized in the "Patients with cardiovascular diseases only" group. The main diseases indicated by these patients in the questionnaire were high blood pressure (232/403, 57.6%), diabetes (120/403, 29.8%), and myocardial infarction (100/403, 24.8%).

Patients who answered being both a "patient with cardiovascular disease or a cardiovascular risk factor" and a "patient with a psychological disorder" in the questionnaire were categorized in the "Patients with both cardiovascular and mental disorders" group. The main diseases indicated by these patients in the questionnaire were high blood pressure (302/528, 57.2%), diabetes (170/528, 32.2%), and myocardial infarction (112/528, 21.2%). Patients who answered being a "patient with a psychological disorder" only in the questionnaire were categorized in the "Patients with mental disorders only" group. The main diseases indicated by these patients in the questionnaire were bipolar disorder (145/292, 49.7%), depression (135/292, 46.2%), and anxiety (125/292, 42.8%). In the risk level groups, the risk level of each patient was based on 5 determined risk factors and estimated by assigning 1 point per risk factor. Actual adherence was assessed based on the response to a set of 4 questions. The adherence score was then calculated by assigning points according to patients' answers. Patients were then grouped into 4 compliance categories based on the total points. Profiles were identified by using a 3-class ascending hierarchical classification model based on 12 variables.

The study group consisted of 542/820 (66.1%) women and 278/820 (33.9%) men, with a sex ratio (male/female) of 0.5. Patients ranged in age from 18 to 93 years (mean 55.2 years, SD 12.7 years). The demographic characteristics of the patients are summarized in Table 1.

Regarding disease category, 49.2% (403/820) of patients had cardiovascular disease; the most common conditions reported were high blood pressure (302/528, 57.2%), diabetes (170/528, 32.2%), myocardial infarction (112/528, 21.2%), hypercholesterolemia (100/528, 18.9%), and arrhythmia (92/528, 17.4%). In total, 35.6% (292/820) of patients had a mental disorder; the most common conditions reported were depression (199/417, 47.7%), anxiety (193/417, 46.3%), and bipolar disorder (184/417, 44.1%). A total of 15.2% (125/820) of patients had both conditions. The medical characteristics of the patients are summarized in Table 2.

Figure 1. Study population flowchart from the screening of patients to the grouping according to risk level, medication adherence, and behavior profiles.
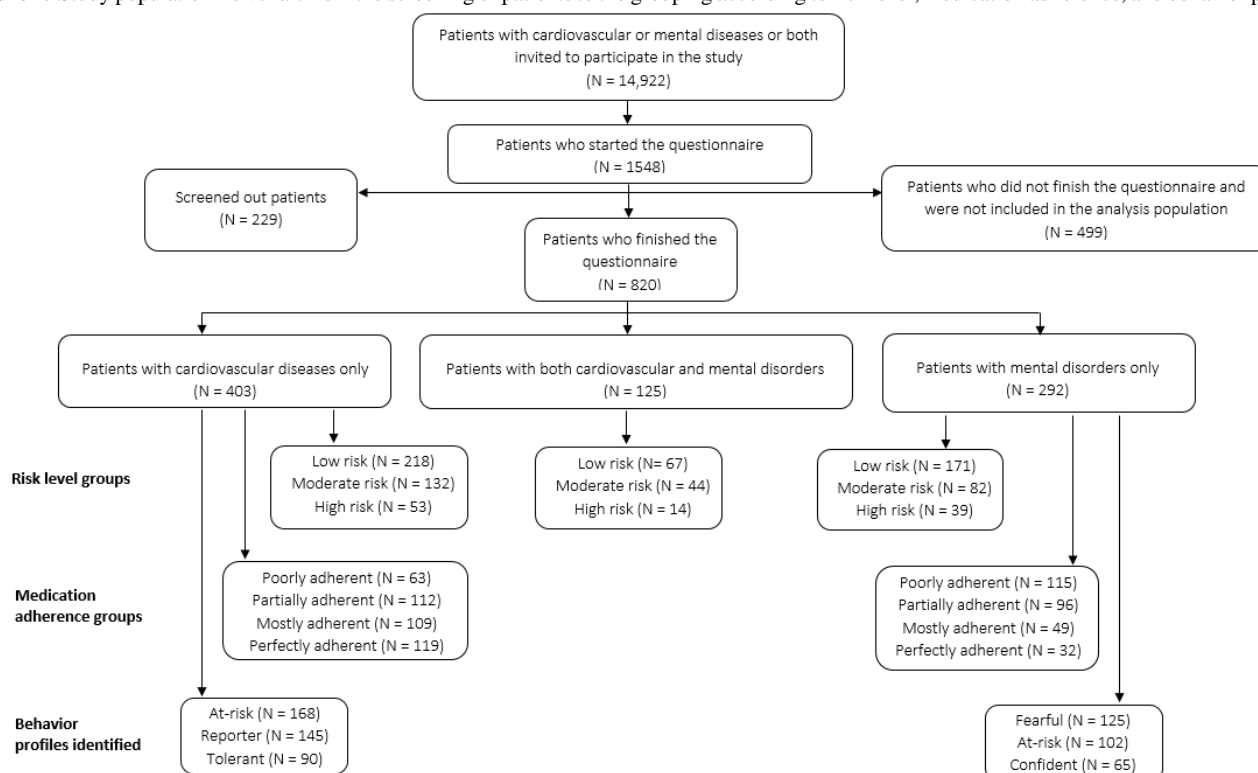
Table 1. Demographic characteristics of the patients with mental disorders or cardiovascular diseases or both conditions who were recruited for the web-based survey on the French Carenity platform.

Demographic characteristics	Total (N=820)	Patients with cardiovascular diseases only (n=403)	Patients with mental disorders only (n=292)	Patients with both conditions (n=125)	P value
Gender, n (%)					≤.001
Female	542 (66.1)	217 (53.8)	228 (78.1)	97 (77.6)	
Male	278 (33.9)	186 (46.2)	64 (21.9)	28 (22.4)	
Age group (years), n (%)					≤.001
≥65	189 (23)	149 (37)	19 (6.5)	21 (16.8)	
<65	631 (77)	254 (63)	273 (93.5)	104 (83.2)	
Mean age (SD) (years)	55.2 (12.7)	60.6 (11)	47.6 (12.1)	55.9 (10.2)	N/A[a]
Level of education completed, n (%)					.30
None	46 (5.6)	27 (6.7)	16 (5.5)	3 (2.4)	
Elementary to middle school	71 (8.7)	38 (9.4)	21 (7.2)	12 (9.6)	
High school	391 (47.7)	196 (48.7)	133 (45.5)	62 (49.6)	
University	312 (38)	142 (35.2)	122 (41.8)	48 (38.4)	
Professional status, n (%)					≤.001
Currently working	309 (37.7)	142 (35.2)	135 (46.2)	32 (25.6)	
Retired	247 (30.1)	179 (44.4)	26 (8.9)	42 (33.6)	
Other or not active	264 (32.2)	82 (20.4)	131 (44.9)	51 (40.8)	

[a]N/A: not applicable.

Table 2. Medical profile of the patients with mental disorders or cardiovascular diseases or both conditions who were recruited for the web-based survey on the French Carenity platform.

Medical characteristics	Total (N=820)	Patients with cardiovascular diseases only (n=403)	Patients with mental disorders only (n=292)	Patients with both conditions (n=125)	P value[a]
Time since diagnosis (years), n (%)					.20
0-5	193 (23.5)	113 (28)	67 (23)	13 (10.4)	
5-20	330 (40.2)	166 (41.2)	118 (40.4)	46 (36.8)	
>20	162 (19.8)	62 (15.4)	52 (17.8)	48 (38.4)	
Do not remember	135 (16.5)	62 (15.4)	55 (18.8)	18 (14.4)	
Mean (SD) (years)	14.2 (11.7)	12.5 (11)	13.5 (10.6)	21.3 (13.5)	N/A[b]
Number of medications per day, n (%)					≤.001
1-2	263 (32.1)	113 (28)	130 (44.5)	20 (16)	
3-5	312 (38)	158 (39.2)	112 (38.4)	42 (33.6)	
≥6	245 (29.9)	132 (32.8)	50 (17.1)	63 (50.4)	
Mean (SD) number of medications per day	4.7 (3.71)	5.1 (4.17)	3.6 (2.7)	5.9 (3.6)	N/A
Number of pills per day, n (%)					.08
1	79 (9.6)	48 (11.9)	29 (9.9)	2 (1.6)	
2-5	359 (43.8)	181 (44.9)	138 (47.3)	40 (32.)	
≥6	380 (46.4)	173 (42.9)	125 (42.8)	82 (65.6)	
Do not know	2 (0.2)	1 (0.3)	0 (0)	1 (0.8)	
Mean (SD) number of pills per day	6.4 (5.1)	6.1 (5.1)	5.9 (4.6)	8.8 (5.8)	N/A
Complexity of treatment regimen, n (%)					.003
Simple	269 (32.8)	148 (36.7)	97 (33.2)	24 (19.2)	
Intermediate	204 (24.9)	104 (25.8)	66 (22.6)	34 (27.2)	
Complex	347 (42.3)	151 (37.5)	129 (44.2)	67 (53.6)	
Risk level (lifestyle), n (%)					.40
Low risk	456 (55.6)	218 (54.1)	171 (58.5)	67 (53.6)	
Moderate risk	258 (31.5)	132 (32.7)	82 (28.1)	44 (35.2)	
High risk	106 (12.9)	53 (13.2)	39 (13.4)	14 (11.2)	

[a]Cardiovascular diseases vs mental disorders.

[b]N/A: not applicable.

The majority of patients were polymedicated; only 14.1% (116/820) of patients were taking only 1 medication. The average number of medications per day was 4.7 (SD 3.7), with a mean of 6.4 (SD 5.1) pills per day per patient (Table 2). Thus, treatments were very burdensome for some patients who took ≥10 pills (168/820, 20.5%).

Moreover, patients >65 years of age were more often polymedicated and took >3 medications per day (116/189, 61.4%) with an average of 5.1 (SD 3.3) medications per day.

Lifestyle and Risk Level

At the time of the survey, a quarter of patients (210/820, 25.6%) were smokers. Most patients (605/820, 73.8%) consumed alcohol, albeit at frequencies that varied between "a few times per year" (278/820, 33.9%) to "at least 3 drinks per day" (20/820, 2.4%).

The mean BMI was 29 (SD 7) kg/m^2, and 69.1% (567/820) of participants were overweight or obese (BMI ≥ 25). Most patients (582/820, 71%) indicated that they did not adhere to any type of diet.

Iatrogenesis

Regarding iatrogenic risk, 84.7% (695/820) of patients indicated worrying about risks associated with their medication; the main reasons were fear of dependence on treatment (372/695, 53.5%), lack of medication efficacy (373/695, 53.7%), the effect the medication will have on another condition (304/695, 43.7%), and intolerance or allergy to the medication (291/695, 41.9%). Less adherent patients are more likely to worry about dependency on treatment than perfectly adherent patients (89/199, 44.7% vs 36/128, 28.1%; P=.004). Nearly half of patients (432/820, 52.7%) worried about treatment risks and constraints when they started a new treatment; 73.1% (316/432) of these patients anticipated adverse effects.

A large majority of patients (722/820, 88%) worried about at least one treatment risk or constraint. These patients had discussed these issues with their general practitioner (307/722, 42.5%), a health specialist (305/722, 42.2%), or a relative (141/722, 19.5%).

Regarding adverse effects, 82.8% (679/820) of patients experienced adverse effects, and 19.6% (133/679) of these patients disrupted their treatment because of these effects without consulting with their doctors.

In terms of the information on the medications, 73% (599/820) of patients reported having received information from their doctors about treatment benefits, but only 44.6% (366/820) of patients had been informed about possible adverse effects for each medication. Patients with cardiovascular diseases tend to receive explanations on the benefit and health impacts of all their treatments more often than patients with mental disorders (317/403, 78.6% vs 195/292, 66.8%; $P \leq .001$). Furthermore, 35.8% (294/820) of patients had only received partial information regarding the benefits for health and potential interactions and adverse effects associated with their medication. Regarding specific medications, 52.4% (154/294) of patients were informed about most of their medications, 47.6% (140/294) of patients received information about some medications, and 20% (160/820) of patients did not receive any information. Moreover, 26.2% (215/820) of patients were unaware of possible pharmaceutical interactions, and only 40.1% (329/820) of patients were aware of the risks of potential interaction between medications.

Treatment Adherence

Perceived and Observed Treatment Adherence

In this study, the criteria for perfect medication adherence included taking all medication on time and as prescribed, without any interruptions. Overall, patients tended to perceive themselves to be more adherent than they actually were; 70.7% (580/820) of patients perceived themselves to be highly adherent (mean 8.9, median 9.9), 66.3% (544/820) reported taking their medication late or early, and 26.4% (217/820) regularly disregarded the prescribed dose. Patients with cardiovascular diseases perceived themselves to be highly adherent more often than those with mental disorders (320/403, 79.4% vs 173/292, 59.2%; $P \leq .001$). Furthermore, many patients stopped or interrupted a medication either unintentionally (429/820, 52.3%)

or intentionally (254/820, 31%) without consulting with a doctor. The most common reasons provided by the 646 patients who did not take their medication as prescribed were that they had forgotten to take it (335/646, 51.9%) or that they wanted to avoid experiencing adverse effects (188/646, 29.1%). Patients with mental disorders were more likely to intentionally disrupt their treatment than those with cardiovascular diseases (151/260, 58.1% vs 97/284, 34.1%; $P \leq .001$), primarily because they wanted to avoid experiencing adverse effects (84/260, 32.3%). When patients with cardiovascular diseases intentionally disrupted their treatment, they also most often mentioned that they did so to avoid experiencing adverse effects (63/284, 22.2%).

Factors Associated With Treatment Adherence

Factors associated with treatment adherence included demographic characteristics, disease type, treatment characteristics, and patient awareness of treatment and the associated adverse effects (Table 3). Men were more adherent than women (152/278, 54.7% vs 204/542, 37.6% mostly perfectly adherent patients; $P \leq .001$). The older patients (≥ 65 years) were markedly more adherent than the younger patients (112/189, 59.3% vs 244/631, 38.7% mostly perfectly adherent patients; $P \leq .001$). The more complex the treatment regimen (eg, multiple treatments during the day, frequent dose changes), the less adherent the patients were to the treatment regimen (mostly perfectly adherent patients: 75/269, 27.9% simple treatment vs 62/347, 17.9% complex treatment; $P \leq .001$). We also found that patients with cardiovascular diseases were more adherent than patients with mental disorders (228/403, 56.6% vs 81/292, 27.7% mostly perfectly adherent patients; $P \leq .001$).

Patients who were less adherent were more likely to think that they would experience adverse effects when they started a new medication (126/216, 58.3% of poorly adherent patients vs 70/174, 40.2% of perfectly adherent patients; $P \leq .001$). Similarly, the less adherent patients were to the medication, the more likely they were to worry about treatment constraints (160/216, 74.1% of poorly adherent patients vs 90/174, 51.7% of perfectly adherent patients; $P \leq .001$). We also found that poorly adherent patients were less likely to receive information regarding treatment benefits and adverse effects for each medication (74/216, 34.2% of poorly adherent patients vs 102/174, 58.6% of perfectly adherent patients; $P \leq .001$; Table 3).

Table 3. Factors associated with treatment adherence according to adherence level among patients with mental disorders or cardiovascular diseases or both conditions who were recruited for the French Carenity platform study (N=820).

Factors	Total	Poorly adherent, n (%)	Partially adherent, n (%)	Mostly adherent, n (%)	Perfectly adherent, n (%)	P value
Gender						≤.001
Female	542	166 (30.6)	172 (31.7)	100 (18.5)	104 (19.2)	
Male	278	50 (18)	76 (27.3)	82 (29.5)	70 (25.2)	
Age group (years)						≤.001
≥65	189	27 (14.3)	50 (26.5)	46 (24.3)	66 (34.9)	
<65	631	189 (30)	198 (31.4)	136 (21.6)	108 (17.1)	
Disease type						≤.001
Cardiovascular diseases only	403	63 (15.6)	112 (27.8)	109 (27)	119 (29.5)	
Mental disorders only	292	115 (39.4)	96 (32.9)	49 (16.8)	32 (11)	
Complexity of treatment regimen						≤.001[a]
Simple	269	58 (21.6)	75 (27.9)	61 (22.7)	75 (27.9)	
Intermediate	204	52 (25.5)	61 (29.9)	54 (26.5)	37 (18.1)	
Complex	347	106 (30.5)	112 (32.3)	67 (19.3)	62 (17.9)	
Fear of treatment risks						≤.001
No	125	17 (13.6)	23 (18.4)	39 (31.2)	46 (36.8)	
Yes	695	199 (28.6)	225 (32.4)	143 (20.6)	128 (18.4)	
Worry about treatment constraints						≤.001
No	262	56 (21.4)	72 (27.5)	80 (30.5)	84 (32.1)	
Yes	528	160 (30.3)	176 (33.3)	102 (19.3)	90 (17.1)	
Received information on treatment benefits and impact on health						≤.001
Yes	599	126 (21)	188 (31.4)	144 (24)	141 (23.6)	
Yes partially	142	64 (45.1)	42 (29.6)	21 (14.8)	15 (10.5)	
No	79	26 (32.9)	18 (22.8)	17 (21.5)	18 (22.8)	
Received information on adverse effects						≤.01
Yes	660	162 (24.5)	208 (31.5)	148 (22.4)	142 (21.5)	
No	160	54 (33.7)	40 (25)	34 (21.3)	32 (20)	

[a]The P value only refers to the comparison of the simple and complex treatment regimens.

Reporting of Adherence Issues

Among the patients who did not always take their medication as prescribed, 59.1% (382/646) did not consult with their physician because they did not think it was necessary (146/382, 38.2%), they forgot (116/382, 30.4%), or they were afraid of disappointing their doctors or of being judged (87/382, 22.8%). Patients with the most risk factors (ie, overweight, smoking, age ≥65 years) seemed least likely to always notify their physician of adherence issues (26/81, 32.1% of high-risk patients vs 162/363, 44.6% of low-risk patients; P=.05).

Solutions and Tools Used by Patients

Just over half of respondents (428/820, 52.2%) used at least one tool to help them properly adhere to their treatment, such as a pill organizer (334/820, 40.7%) or an alarm (80/820, 9.7%). Only 1.9% (16/820) of patients used a smartphone application. Patients with cardiovascular diseases were more inclined to use

a tool than those with mental disorders (224/403, 55.6% vs 136/292, 46.6%; P=.02). A total of 12.2% (100/820) of patients reported that they received help with their medication from a caregiver. Among those who received assistance, the caregiver primarily reminded them to take their medication (60/100, 60%) or purchased the medication for them (43/100, 43%).

Patient Behavior Profiles

Patients With Cardiovascular Diseases Only

Patients with cardiovascular diseases were classified as follows based on 3 identified behavior profiles: at-risk patients, tolerant patients, and reporter patients.

The at-risk patients (168/403, 41.7%) were the least adherent (110/168, 65.5% were not fully adherent). The majority of these patients were women (121/168, 72%). Although these patients perceived themselves to be adherent (123/168, 73.2%), they took ≤3 pills per day (71/168, 42.3%), used complementary

approaches (138/168, 82.1%), and intentionally deviated from initial prescriptions (65/140, 46.4% of these patients did not take their medication as prescribed). These patients were worried about the treatment interfering with their daily routine (55/168, 32.7%) and the frequency of administration (68/168, 40.5%). Moreover, 33.9% (57/168) of the at-risk patients were more likely to worry about becoming dependent on their medication, and 49.4% (83/168) of patients worried about medication efficacy. These patients visited their physician more often (139/168, 82.7%) but did not always notify them of nonadherence (88/168, 52.4%) because of a fear of disappointing them (80/88, 47.86%). Finally, 73.2% (123/168) of these patients did not receive information about medication benefits, and 75% (126/168) of patients had discussed constraints and risks with someone.

The reporter patients (145/403, 36%) were largely adherent (116/145, 80%) and perceived themselves to be adherent (131/145, 90.3% of these patients perceived themselves to be fully adherent). The majority of these patients primarily included older patients (66/145, 45.5% ≥65 years of age) and men (90/145, 62.1%) who took >3 pills per day (118/145, 81.4%). They also believed that they were well-informed (122/145, 84.1%), they used practical tools to manage their medications (91/145, 62.7%), and they were aware of potential adverse interactions between their medication and tobacco (54/145, 37.2%) and alcohol (74/145, 51%).

The tolerant patients (90/403, 22.3%) experienced fewer adverse effects (78/90, 86.7%) and did not adjust their treatment (85/90, 94.4%). These patients were less aware of adverse interactions between their treatment and other medications (66/90, 73.3%) and between their treatment and alcohol (58/90, 64.4%). These patients did not inform their doctors about adherence issues (32/42, 76.2%).

Patients With Mental Disorders Only

Patients with mental disorders were grouped as follows based on 3 identified behavior profiles: fearful patients, at-risk patients, and confident patients.

The fearful patients (125/292, 42.8%) were polymedicated (85/125, 68% of patients took ≥3 pills per day), had experienced treatment adverse effects (125/125, 100%), and anticipated adverse effects when they started a treatment (75/125, 60%). These patients were concerned about treatment risks (101/125, 80.8%) and adjusted their treatment in the event of adverse effects (55/125, 44%); 25.6% (32/125) stopped the medication and 8.8% (11/125) modified the dose. These patients also used fewer complementary approaches such as nonpharmaceutical alternatives (54/125, 43.2%), homeopathy, (16/125, 12.8%), and acupuncture (3/125, 2.4%).

The at-risk patients (102/292, 34.9%) were the least adherent patients (82/102, 80% of these patients were not fully adherent), and the majority of these patients were women (90/102, 88.2%). They took only 1-2 pills per day (57/102, 55.9%) and perceived themselves to be poorly adherent (55/102, 53.9%). These patients worried about treatment risks and constraints (82/102, 80.4%). All patients used at least one other product or an alternative (102/102, 100%); nearly all patients (94/102, 92.1%)

used nonpharmaceutical therapy, 89.2% (91/102) of patients used homeopathy, and 52% (53/102) used acupuncture. They experienced adverse effects (102/102, 100%) and adjusted their treatment in case of an adverse effect (45/102, 44.1%); 17.6% (18/102) stopped the medication and 14.7% (15/102) modified the dose. Of these patients, 75.5% (77/102) reported adverse effects to their physician and 67.6% (69/102) had discussed constraints and risks with their pharmacist.

The confident patients (65/292, 22.3%) were more adherent (27/65, 42%) and the majority of those who were adherent were men (44/65, 67.7%). Few patients experienced adverse effects (44/65, 67.7% did not experience adverse effects); these patients did not adjust their treatment in the event of an adverse effect (60/65, 92.3%) and they were less likely to use complementary approaches (27/65, 41.4%). These patients worried less about the risks associated with their medication (22/65, 33.8%). They were less informed about risks (21/65, 32.3%) but well informed about treatment benefits (50/65, 76.9%). Finally, among the confident patients who did not always notify their doctors in the event of therapeutic nonadherence, 42.8% (15/35) did not think it was necessary to do so.

Discussion

Principal Findings

Web-based communities represent an increasingly popular and accessible platform for patients to learn about their condition and participate in clinical studies [22]. These web-based patient communities also provide researchers access to data about specific patient populations that are demographically representative. These tools enable the assessment of data concerning patient behavior, experiences, and well-being in all aspects of their lives (medical, professional, and personal), which are difficult to observe using other methods [22].

The current study focused on perceptions of iatrogenic risk and treatment adherence in patients with cardiovascular diseases, mental disorders, or both cardiovascular diseases and mental disorders. All patients were registered on the largest web-based Carenity patient community in France. We found that 82.8% (679/820) of patients experienced adverse effects associated with their medication. While the majority of patients (492/679, 72.4%) informed their doctor about adverse effects, many patients took steps to address adverse events on their own, with 30.8% (209/679) disrupting their treatment without medical advice. These results indicate that adverse effects present an understandable challenge for patients [23] and represent a major barrier to medication adherence [24]. We also found that well-informed patients were more likely to report adverse effects to a health care professional and are less likely to disrupt their treatment on their own. The frequent report of adverse events likely explains why most patients anticipated adverse effects when they started a new medication and worried about treatment risks and constraints such as dependence, lack of efficacy, effects on another condition, and intolerance.

Regarding therapeutic adherence, patients tended to perceive themselves to be more adherent than they actually were. Medication adherence is often overestimated by patients [25].

In this study, more than half the number of patients (426/820, 52%) unintentionally stopped or interrupted treatment at least once, often because they forgot or wanted to avoid experiencing adverse effects.

The results are consistent with those of previous studies that have indicated that approximately 50% of patients undergoing long-term therapy were nonadherent to their treatment [9,11]. A separate study has also indicated that forgetting to take medication and having limited awareness were the most frequently cited reasons for treatment nonadherence [14,26]. Previous studies have also shown that several factors are associated with adherence, including patient characteristics (gender, age, ethnicity, marital status), symptom intensity, medication type, route of administration, and severity of adverse effects [27,28]. In this study, the factors associated with treatment adherence included gender, age, treatment complexity, and disease type. Men were more adherent than women, and patients ≥65 years of age were significantly more adherent than the younger patients. Moreover, patients with a simple treatment regimen and those who had received information about treatment benefits and potential adverse effects tended to be more adherent. The degree of adherence also varied between the 2 disease groups, as patients with cardiovascular diseases tended to be more adherent than those with mental disorders. Less adherent patients were less aware of potential drug interactions, although they were more likely to anticipate adverse effects when they started a new treatment. They also worried more about risks of treatment dependency, lack of treatment efficacy, and the constraints that therapy may impose on their daily routine.

It is important to emphasize that the study was performed during the ongoing COVID-19 pandemic. To date, limited research is available on how the pandemic has specifically affected therapeutic adherence in NCD patients, but it is well known that the COVID-19 pandemic has disrupted health care services around the world, which has possible implications on therapeutic adherence in patients. One systematic review reported a significant failure of patients with inflammatory bowel disease to adhere to therapies during the COVID-19 pandemic [29]. However, as shown in a pharmacoepidemiological study conducted using data from the French National Health Data System, the COVID-19 pandemic did not seem to lead to a shortage of treatment for patients with cardiovascular diseases. For example, a significant number of hypertensive patients overstocked their medication when the lockdown was announced. On the other hand, the study demonstrated a decrease in consumption of nonprescription drugs and products needed for examinations, such as colonoscopies, or contrast agents. Therefore, the combination of factors related to social restriction (lockdowns, difficulty in accessing health care or treatments) and patient-related factors (fear of infection, decision to take or not take the drug, treatment dosage adjustment) may have impacted therapeutic adherence in NCD patients, even though this impact seems to be more nuanced in France than in other countries [30].

The 3 patient profiles identified for each disease type provide significant insight for improving therapeutic adherence and adjustment strategies specifically according to patient behaviors.

With respect to the cardiovascular disease profiles, the at-risk patients, who were less adherent and worried about treatment risks and adverse effects, should be better informed about treatment risks (especially adverse effects) and treatment benefits. The tolerant patients, who were polymedicated, more adherent, and less worried about risks and adverse effects, should be better informed about tools they can use to properly manage their medication. The reporter patients, who did not discuss adverse effects with their pharmacists and were unaware of treatment interactions, should be better informed about the risks of treatment interactions and encouraged to notify their doctors when medications are not taken as prescribed.

Regarding the mental disorder profiles, the at-risk patients, who were not adherent, rarely communicated with their doctors, and overestimated treatment risks and constraints, should have access to specialized therapeutic educational programs to improve awareness and medication adherence (eg, shared decision making) [31]. The fearful patients, who were polymedicated, experienced adverse effects, and were better informed about drug interactions and adverse effects, should be closely monitored by their doctors so that advice focused on their specific needs can be provided. Finally, the confident patients, who tended not to seek assistance when needed and did not present major challenges, should be encouraged to build a better relationship with their doctors and seek assistance when necessary.

Limitations

A few study limitations should be mentioned. As the study was based on data collected via a web-based survey, it may exclude patients who are not comfortable using or do not have access to internet or a computer or who are non-French speakers living in France. The underrepresentation of older patients in the Carenity community and the overrepresentation of patients who are actively concerned about their health may have also led to selection bias. Moreover, patient characteristics and medication adherence were assessed using self-reported measures, which may have led to recall bias. Finally, the group of patients with both cardiovascular diseases and mental disorders could not be included in the statistical analysis because of the small number of patients in the group. Patients with multiple NCDs should be included in future studies because they are at an increased risk of iatrogenic disease [18].

Patients completed the questionnaire without the guidance of their physicians, so desirability bias was greatly limited. As individuals registered in patient communities may likely be heavily burdened by their condition, our study population may overrepresent highly symptomatic and polymedicated patients. Nevertheless, it has been shown that characteristics of patients in Carenity communities reflect the main characteristics of web-based users willing to share their medical experience but with an overrepresentation of female patients aged 25-54 years [32]. This study exclusively included patients registered on the French Carenity platform and did not include other relevant sampling procedures and methods. Despite the fact that study recruitment focused on a homogeneous population of patients through the Carenity platform, the results of the study are not

generalizable to the larger population of patients with these disorders.

Conclusions

Overall, the results of this web-based survey study provide important insight into patients' perspectives and behaviors because the anonymous nature of the survey allowed patients to respond openly and honestly. These findings emphasize the importance of involving patients in medical decisions and providing patients with information about treatment benefits, treatment adherence, potential adverse effects and risks of treatment, and potential drug interactions. Therapeutic alliance significantly helps patients to understand both the disease aspect and the therapeutic options, thereby improving medication adherence and overall disease management [11,18]. Integrative and comprehensive patient care that considers the complementary therapeutic approaches used by patients could also improve medication adherence [26]. Our results also show that caregivers and pharmacists should be empowered to proactively support and better educate patients with an NCD who require multiple medications. Practical tools should be developed to remind patients to take their medication as prescribed, and additional studies should assess improved support strategies for patients with chronic diseases.

Classical adherence studies are often biased and do not accurately represent the patient population for a given disease. Web-based platforms on which patients can share their medical experiences in an anonymous manner may provide unique insight into the perspectives of patients undergoing therapy for an NCD, which could help to improve disease management and eventually prevent premature deaths.

Conflicts of Interest

PC is the Head of the Department of Psychiatric Emergency and Acute Care at Centre Hospitalier Universitaire de Montpellier. PC has received fees for occasional consultation for Janssen, Exeltis, and Pfizer. CP is a Medical Director at Viatris and discloses stock holdings from Pfizer and Viatris. AFLP is a Quality Manager at Pfizer PFE France, a Viatris company. CA is an employee of Carenity, and MC is the President and Founder of Carenity; CA and MC have no conflicts of interest to declare. JJM is Head of the Internal Medicine Department at Groupe Hospitalier Paris Saint-Joseph (Paris). JJM has received fees for occasional consultation for Novartis, Servier, Mylan, and Pfizer.

References

1. Noncommunicable diseases. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2021-04-25]

2. Global Action Plan for the Prevention and Control of NCDs 2013-2020. In: World Health Organization. Geneva: WHO; Nov 14, 2013.

3. Global Action Plan for the Prevention and Control of NCDs 2013-2020. In: World Health Organization. Geneva: WHO; 2014:280.

4. The WHO Special Initiative for Mental Health (2019-2023): Universal Health Coverage for Mental Health. In: World Health Organization. Geneva: WHO; May 2, 2019.

5. Nilson EAF, Metlzer AB, Labonté ME, Jaime PC. Modelling the effect of compliance with WHO salt recommendations on cardiovascular disease mortality and costs in Brazil. PLoS One 2020 Jul 9;15(7):e0235514 [FREE Full text] [doi: 10.1371/journal.pone.0235514] [Medline: 32645031]

6. Mensah GA, Collins PY. Understanding mental health for the prevention and control of cardiovascular diseases. Glob Heart 2015 Sep 01;10(3):221-224 [FREE Full text] [doi: 10.1016/j.gheart.2015.08.003] [Medline: 26407518]

7. Miller TA. Health literacy and adherence to medical treatment in chronic and acute illness: a meta-analysis. Patient Educ Couns 2016 Jul;99(7):1079-1086 [FREE Full text] [doi: 10.1016/j.pec.2016.01.020] [Medline: 26899632]

8. Cramer JA, Roy A, Burrell A, Fairchild CJ, Fuldeore MJ, Ollendorf DA, et al. Medication compliance and persistence: terminology and definitions. Value Health 2008;11(1):44-47 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00213.x] [Medline: 18237359]

9. Kini V, Ho PM. Interventions to improve medication adherence: a review. JAMA 2018 Dec 18;320(23):2461-2473. [doi: 10.1001/jama.2018.19271] [Medline: 30561486]

10. Zagożdżon P, Wrotkowska M. Religious beliefs and their relevance for treatment adherence in mental illness: a review. Religions 2017 Aug 14;8(8):150. [doi: 10.3390/rel8080150]

11. Brown MT, Bussell JK. Medication adherence: WHO cares? Mayo Clin Proc 2011 Apr;86(4):304-314 [FREE Full text] [doi: 10.4065/mcp.2010.0575] [Medline: 21389250]

12. U.S. Food and Drug Administration. Medication Adherence: Landscape, Strategies, and Evaluation Methods. 2019 Presented at: Public Workshop held by FDA and Duke-Margolis; December 10, 2019; Washington, DC URL: https://www.fda.gov/drugs/news-events-human-drugs/medication-adherence-landscape-strategies-and-evaluation-methods-12102019

13. Jack K, McLean SM, Moffett JK, Gardiner E. Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review. Man Ther 2010 Jun;15(3):220-228 [FREE Full text] [doi: 10.1016/j.math.2009.12.004] [Medline: 20163979]

14. Aminde LN, Tindong M, Ngwasiri CA, Aminde JA, Njim T, Fondong AA, et al. Adherence to antidiabetic medication and factors associated with non-adherence among patients with type-2 diabetes mellitus in two regional hospitals in Cameroon. BMC Endocr Disord 2019 Apr 03;19(1):35 [FREE Full text] [doi: 10.1186/s12902-019-0360-9] [Medline: 30943969]

15. Gonzalez J, Williams JW, Noël PH, Lee S. Adherence to mental health treatment in a primary care clinic. J Am Board Fam Pract 2005 Mar 01;18(2):87-96 [FREE Full text] [doi: 10.3122/jabfm.18.2.87] [Medline: 15798137]

16. Barnard-Kelly KD, Naranjo D, Majidi S, Akturk HK, Breton M, Courtet P, et al. An intolerable burden: suicide, intended self-injury and diabetes. Can J Diabetes 2020 Aug;44(6):541-544 [FREE Full text] [doi: 10.1016/j.jcjd.2020.01.008] [Medline: 32305294]

17. Mercier E, Giraudeau B, Giniès G, Perrotin D, Dequin P. Iatrogenic events contributing to ICU admission: a prospective study. Intensive Care Med 2010 Jun 9;36(6):1033-1037. [doi: 10.1007/s00134-010-1793-9] [Medline: 20217046]

18. Permpongkosol S. Iatrogenic disease in the elderly: risk factors, consequences, and prevention. CIA 2011 Mar:77. [doi: 10.2147/cia.s10252]

19. Ravoire S, Lang M, Perrin E, participants of Giens XXXII Round Table No. 6. Advantages and limitations of online communities of patients for research on health products. Therapie 2017 Feb;72(1):135-143. [doi: 10.1016/j.therap.2016.11.058] [Medline: 28139283]

20. Health Forum: Share with patients and their carers at Carenity. Carenity. URL: https://www.carenity.co.uk/ [accessed 2021-07-01]

21. Body mass index – BMI. World Health Organization. URL: https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi [accessed 2021-05-13]

22. Patel AT, Wein T, Bahroo LB, Wilczynski O, Rios CD, Murie-Fernández M. Perspective of an international online patient and caregiver community on the burden of spasticity and impact of botulinum neurotoxin therapy: survey study. JMIR Public Health Surveill 2020 Dec 07;6(4):e17928 [FREE Full text] [doi: 10.2196/17928] [Medline: 33284124]

23. Thompson L, McCabe R. The effect of clinician-patient alliance and communication on treatment adherence in mental health care: a systematic review. BMC Psychiatry 2012 Jul 24;12(1). [doi: 10.1186/1471-244x-12-87]

24. Kleinsinger F. The unmet challenge of medication nonadherence. Perm J 2018;22:18-033. [doi: 10.7812/TPP/18-033]

25. Stentzel U, van den Berg N, Schulze LN, Schwaneberg T, Radicke F, Langosch JM, et al. Predictors of medication adherence among patients with severe psychiatric disorders: findings from the baseline assessment of a randomized controlled trial (Tecla). BMC Psychiatry 2018 May 29;18(1):155 [FREE Full text] [doi: 10.1186/s12888-018-1737-4] [Medline: 29843676]

26. Kvarnström K, Airaksinen M, Liira H. Barriers and facilitators to medication adherence: a qualitative study with general practitioners. BMJ Open 2018 Jan 23;8(1):e015332 [FREE Full text] [doi: 10.1136/bmjopen-2016-015332] [Medline: 29362241]

27. Gellad WF, Grenard JL, Marcum ZA. A systematic review of barriers to medication adherence in the elderly: looking beyond cost and regimen complexity. Am J Geriatr Pharmacother 2011 Feb;9(1):11-23 [FREE Full text] [doi: 10.1016/j.amjopharm.2011.02.004] [Medline: 21459305]

28. Jimmy B, Jose J. Patient medication adherence: measures in daily practice. Oman Med J 2011 May;26(3):155-159 [FREE Full text] [doi: 10.5001/omj.2011.38] [Medline: 22043406]

29. Jena A, Singh AK, Kumar- M, Sharma V, Sebastian S. Systematic review on failure to adhere to IBD therapies during the COVID-19 pandemic: correct information is crucial. Dig Liver Dis 2020 Nov;52(11):1254-1256 [FREE Full text] [doi: 10.1016/j.dld.2020.09.012] [Medline: 33011089]

30. Dray-Spira R, Drouin J, Le Tri T, Zureik M, Weill A. Évolution de l'exposition au valproate pendant la grossesse et parmi les femmes en âge de procréer en France depuis 2015 : estimations à partir des données du Système national des données de santé. Revue d'Épidémiologie et de Santé Publique 2020 Mar;68:S4. [doi: 10.1016/j.respe.2020.01.005]

31. Samalin L, Genty J, Boyer L, Lopez-Castroman J, Abbar M, Llorca P. Shared decision-making: a systematic review focusing on mood disorders. Curr Psychiatry Rep 2018 Mar 27;20(4):23. [doi: 10.1007/s11920-018-0892-0] [Medline: 29589129]

32. Raïs S, Radoszycki L, Dourgnon P, Rochaix L, Chekroun M. Accurate representation of patients' opinions for decision-making: are online health communities good candidates? Value Health 2017 Oct;20(9):A760-A761. [doi: 10.1016/j.jval.2017.08.2153]

Abbreviations

NCD: noncommunicable disease

XSL•FO
RenderX

Original Paper

Impact of a Conformité Européenne (CE) Certification–Marked Medical Software Sensor on COVID-19 Pandemic Progression Prediction: Register-Based Study Using Machine Learning Methods

Leevi Limingoja[1], BM; Kari Antila[2], PhD; Vesa Jormanainen[1,3], MD, MSc; Joel Röntynen[4], PhD; Vilma Jägerroos[4], MSc; Leena Soininen[5], MD, EMBA, PhD; Hanna Nordlund[6], CSC; Kristian Vepsäläinen[3,7], MSc; Risto Kaikkonen[4], MSoc; Tea Lallukka[1], PhD

[1]Department of Public Health, University of Helsinki, Helsinki, Finland

[2]Solita Oy, Tampere, Finland

[3]Health and Social Service System Research, Welfare State Research and Reform, Finnish Institute for Health and Welfare, Helsinki, Finland

[4]Solita Oy, Helsinki, Finland

[5]Digifinland Oy, Helsinki, Finland

[6]IT Centre for Science, Helsinki, Finland

[7]University of Eastern Finland, Kuopio, Finland

Corresponding Author:
Leevi Limingoja, BM
Department of Public Health
University of Helsinki
PO Box 20 (Tukholmankatu 8 B)
Helsinki, FIN-00014
Finland
Phone: 358 400344644
Email: leevi.limingoja@helsinki.fi

Abstract

Background: To address the current COVID-19 and any future pandemic, we need robust, real-time, and population-scale collection and analysis of data. Rapid and comprehensive knowledge on the trends in reported symptoms in populations provides an earlier window into the progression of viral spread, and helps to predict the needs and timing of professional health care.

Objective: The objective of this study was to use a Conformité Européenne (CE)-marked medical online symptom checker service, Omaolo, and validate the data against the national demand for COVID-19–related care to predict the pandemic progression in Finland.

Methods: Our data comprised real-time Omaolo COVID-19 symptom checker responses (414,477 in total) and daily admission counts in nationwide inpatient and outpatient registers provided by the Finnish Institute for Health and Welfare from March 16 to June 15, 2020 (the first wave of the pandemic in Finland). The symptom checker responses provide self-triage information input to a medically qualified algorithm that produces a personalized probability of having COVID-19, and provides graded recommendations for further actions. We trained linear regression and extreme gradient boosting (XGBoost) models together with F-score and mutual information feature preselectors to predict the admissions once a week, 1 week in advance.

Results: Our models reached a mean absolute percentage error between 24.2% and 36.4% in predicting the national daily patient admissions. The best result was achieved by combining both Omaolo and historical patient admission counts. Our best predictor was linear regression with mutual information as the feature preselector.

Conclusions: Accurate short-term predictions of COVID-19 patient admissions can be made, and both symptom check questionnaires and daily admissions data contribute to the accuracy of the predictions. Thus, symptom checkers can be used to estimate the progression of the pandemic, which can be considered when predicting the health care burden in a future pandemic.

Introduction

Background

The rapid spread of the SARS-CoV-2 virus leading to a pandemic presented challenges for nationwide assessment of the progression of the COVID-19 pandemic [1]. The virus was first discovered in Wuhan, China, in December 2019, and COVID-19 was declared a pandemic by the World Health Organization in March 2020 [2-4]. In Finland, cases started to appear in late February 2020, and the Finnish government announced a national lockdown in mid-March 2020 to slow the viral spread and protect risk groups [1].

Digital health technology tools such as symptom checkers have been used in different countries (eg, Finland, France, Israel, Italy, the Netherlands, the United Kingdom, and the United States) as self-triage tools for possible SARS-CoV-2 infections [5-11]. In Finland, a COVID-19 symptom checker was added to a preexisting national Conformité Européenne (CE)-marked medical symptom checker service, Omaolo [5-12]. The web-based symptom checker provides the user with advice on further actions based on a medically approved algorithm. Although there have been studies about how well symptom checkers perform as clinical tools [13], to our knowledge, the potential of these data for predicting epidemic progression has not yet been studied. Having real-time comprehensive data on reported symptom trends could provide an earlier window into the viral spread and help predict the burden of professional health care.

To study if the data collected by the Omaolo service and the national care notification registers could be used to predict pandemic progression in Finland, we used the methods of machine learning. Perhaps the best potential of machine learning over more traditional methods lies in its ability to better adapt to the data, and thus to the evolution of the underlying phenomenon. With large data sets such as the Omaolo COVID-19 symptom checker responses, machine learning may also uncover more complex associations between the factors contributing to the predicted outcomes. Machine learning models can also be trained and retrained along the way to reveal how the significance of the individual input variables for making the predictions will change over time and to become a more accurate predictor as more data are collected.

Objectives

The study objective was to assess if a nationwide symptom checker can be used as a predictive tool in estimating the national progression of the COVID-19 pandemic and health care admissions by utilizing machine learning methods.

Methods

Data

Omaolo

The COVID-19 epidemic in Finland started in mid-March 2020. On March 16, 2020, the Finnish government announced a state of emergency due to the COVID-19 epidemic, and consequently implemented several physical distancing measures aimed at slowing the spread and protecting risk groups [1]. Part of the national response was the Omaolo COVID-19 web-based symptom self-assessment tool, a CE-marked medical device [5,12]. Omaolo was launched for use March 16, 2020, and was published in the two national languages (Finnish and Swedish) and later also in English. The COVID-19 symptom checker functioned as any other symptom checker in Omaolo, and was jointly developed by DigiFinland Oy, Duodecim Publishing Company Ltd, the Finnish Institute for Health and Welfare (THL), Solita Oy, and Mediconsult Oy.

In the symptom checker, the user answers a set of predefined, expert-validated questions. As a result, it returns self-triage information on how to proceed with one's situation. The progress of filling in the questionnaire from start to finish is recorded to the log files of the service. The respondent has a choice to answer anonymously without including one's personal information in the process. The questionnaire itself includes several background questions such as age, postal code, gender, and reason for filling in the questionnaire; existing medical conditions; whether the respondent has had close contact with a COVID-19–positive person; whether the respondent or close contacts have been ordered to quarantine by a physician; where the respondent thinks they may have caught the virus; and what kind of work the respondent does in regard to contacts with others (the full Omaolo questionnaire is provided in Multimedia Appendix 1). The questionnaire has been updated several times during the pandemic to better coincide with the latest COVID-19 research [5].

During the study period, a total of 547,428 responses were submitted to Omaolo. Of these, the contents of 132,951 responses were unsaved due to technical reasons. Almost all of the unsaved responses were submitted prior to March 28, 2020, when Omaolo was yet not configured to save the anonymous responses. A small number of the anonymous responses were not saved during short maintenance breaks throughout the period. Accounting for these losses, a total of 414,477 responses were available for analyses. Care reminders, the self-triaged recommendations for care as given by the Duodecim Evidence-Based Medicine Electronic Decision Support service [5], were available for all submitted responses, including the unsaved responses. The data were pseudoanonymized prior to the analyses.

National Registers: Hilmo, Avohilmo, and Paavo

We used the established national care notification registries Hilmo and Avohilmo [14] of the THL to estimate the demand of COVID-19–related care. These registers contain structured inpatient (Hilmo) and outpatient (Avohilmo) records from all public and private specialist care hospitals in Finland. These records were combined in the data preprocessing stage and are hereafter referred to collectively as "Hilmo." The data were anonymized of all identifiers before use.

As a supplementary source, we used the publicly available version of the Paavo register maintained by Statistics Finland [15]. Among other variables, Paavo contains basic demographics of Finnish citizens based on the postal code of their residence. We used these data to identify and rectify the regional bias in age distribution of the Omaolo users. These data were anonymous at the source.

Predicting the Daily Use of Health Care Resources

We chose our study period to be from March 16, 2020, corresponding to the release of the COVID-19 symptom checker, to June 15, 2020, as the approximate beginning of a period of low activity in the pandemic following the first wave. For the predictions, we used two regressors: linear regression [16] and extreme gradient boosting (XGBoost) regression [17]. The reasoning behind selecting two regressors was to compare a simple and traditional method (linear regression) to a modern option (XGBoost regression) with many hyperparameters that can be learned from the data. Both methods were implemented with three feature preselection strategies: a human expert (KA), F-score [16], and mutual information [18]. All regressor feature selector combinations were tested separately resulting in six different machine learning models.

We chose a scenario where the number of daily COVID-19–related health care admissions, as extracted from the Hilmo register, was predicted 1 week ahead, every week on Wednesdays. This follows a hypothetical scenario where the resources for the following 7 days would be decided midweek (on Wednesday) to give 2 full days to prepare for the weekend, for example by reassessing the need for extra resources and personnel.

For training, testing, and validation of the models, we used time-series nested cross-validation [19]. This strategy was chosen to ensure that the model is trained and tested with samples independent from the validation set; thus, no information from the samples past the prediction point was used. During cross-validation, the set of features (and other hyperparameters in the case of XGBoost), with which the regressor best generalizes its predictions to unseen data in terms of average prediction error on different validation sets, is selected for a given regressor.

We also chose to train the classifiers with Hilmo and Omaolo data first separately and then combined. This was to test how much, if any, the results would improve if the data from both sources were used.

Profiling the Motives of Omaolo COVID-19 Symptom Checker Users

The Omaolo COVID-19 symptom checker achieved considerable popularity immediately after its release. Tens of thousands of responses per day were submitted during the first week. The submission activity showed clear peaks during infomercials and other major media mentions. To distinguish the users that were truly suspecting a COVID-19 exposure from the users that were visiting Omaolo just out of curiosity, a question about this matter was added to the questionnaire in the form of a simple tick-box on March 28, 2020. After this update, it was found that approximately 40% of the responses had the out-of-curiosity option checked. We then investigated whether it was possible to distinguish the two response profiles (out-of-curiosity or not) and which questions were the best predictors of this behavior. We used a naive Bayes classifier [20], logistic regression [16], and XGBoost binary classifier [17]. We chose naive Bayes and logistic regression because they are widely used in medicine and other fields. XGBoost was included since it is based on a different approach (an ensemble of trees) and thus provides an interesting comparison to the two established methods.

All three models were tested with 5-fold random cross-validation, and the sensitivity and specificity for each fold were computed and finally averaged over all folds. The size of the majority set (not out-of-curiosity) was balanced by undersampling to the size of the minority prior to cross-validation.

All analyses, both the predictions and profiling, were performed using Python version 3.6 [21] with the feature selectors, classifiers, and regressors from Scikit-learn module version 1.0 [22].

Ethical Consideration

Our study data are based on statistical register data at the national level. These register data contain no personalizing identifiers. Therefore, this study does not fall under the purview of laws regarding medical research. The study protocol does not violate any ethical considerations or standards, according to a statement from the Medical Ethics Committee of the Hospital District of Helsinki and Uusimaa in Finland (June 2013).

Results

Predicting the Daily Use of Health Care Resources

Women used the web-based symptom checker more often than men; approximately two-thirds of the filled-in forms were completed by women (Table 1). People of working age were also using the symptom checker more than other age groups. Most of the questionnaires (approximately half) were completed in Southern Finland. Cough was the most common symptom, followed by sore throat, fever, headache, and difficulty breathing.

Table 1. Distribution of age, employment status, symptoms, and region of the filled-in Omaolo questionnaires.

Characteristic	Total, n (%)	Men, n (%)	Women, n (%)
Age (years)			
0-9	8652 (2.08)	4291 (2.81)	4361 (1.66)
10-19	30,805 (7.41)	10,689 (6.99)	20,116 (7.68)
20-29	93,511 (22.50)	30,241 (19.77)	63,270 (24.10)
30-39	98,041 (23.59)	34,727 (22.71)	63,314 (24.11)
40-49	77,869 (18.74)	29,010 (18.97)	48,859 (18.61)
50-59	59,432 (14.30)	22,788 (14.90)	36,644 (13.96)
60-69	31,935 (7.69)	13,392 (8.76)	18,543 (7.06)
70-79	12,824 (3.09)	6519 (4.26)	6305 (2.40)
≥80	2453 (0.59)	1286 (0.84)	1167 (0.44)
Total	415,522	152,943	262,579
Employment status			
Not currently working	133,199 (32.59)	47,674 (31.90)	85,525 (32.99)
Health care worker	65,379 (16.00)	7953 (5.32)	57,426 (22.15)
Cannot avoid contact (service worker)	102,544 (25.09)	41,948 (28.07)	60,596 (23.37)
Can avoid contact	107,558 (26.32)	51,869 (34.71)	55,689 (21.48)
Total	408,680	149,444	259,236
Symptoms			
Cough	140,661 (34.60)	53,364 (36.53)	87,297 (33.51)
Trouble breathing	31,143 (7.66)	11,312 (7.74)	19,831 (7.61)
Sore throat	86,985 (21)	24,505 (16.78)	62,480 (23.99)
Headache	35,886 (8.83)	12,233 (8.37)	23,653 (9.08)
Myalgia	10,213 (2.51)	4552 (3.12)	5661 (2.17)
Vomiting	7406 (1.82)	2640 (1.81)	4766 (1.83)
Fever	59,041 (14.52)	23,842 (16.32)	35,199 (13.51)
Loss of smell	433 (0.11)	182 (0.12)	251 (0.10)
Dysphagia	7971 (1.96)	2946 (2.02)	5025 (1.93)
Trismus	6469 (1.59)	1938 (1.33)	4531 (1.74)
Trouble speaking	6392 (1.57)	2971 (2.03)	3421 (1.31)
Other	13,952 (3.43)	5589 (3.83)	8363 (3.21)
Total	406,552	146,074	260,478
Region			
Western central Finland (city of Tampere)	42,759 (14.50)	15,370 (14.41)	27,389 (14.55)
Western coastal Finland (city of Turku)	36,321 (12.31)	13,155 (12.33)	23,166 (12.30)
Northern Finland (city of Oulu)	27,419 (9.30)	9894 (9.27)	17,525 (9.31)
Southern Finland (city of Helsinki)	149,145 (50.56)	54,447 (51.04)	94,698 (50.29)
Eastern Finland (city of Kuopio)	39,340 (13.34)	13,819 (12.95)	25,521 (13.55)
Total	294,984	106,685	188,299

For the analyses, we chose two regressors, linear regression and XGBoost, with three feature preselection strategies for each, and compared their performance. To predict the COVID-19–related admissions for each day, 7 days ahead of the prediction point, the features given to the model were extracted from the responses and the Hilmo register on the 7 and 14 days prior (lag variables). The use of lag variables essentially means that two sets of the time-dependent features

were formed, with the first delayed 7 days and the second delayed 14 days. This was to ensure that no data from any of the sources, Omaolo or Hilmo, were leaked past the point of prediction during feature selection, model training, or testing. The regressors were first trained with 5 expert-selected features: how many of the questionnaires were filled-in by people over 60 years old, how many reported lengths of symptoms were greater than 10 days, and how many were assigned the urgency code P1 (the most urgent) in the care recommendations, in addition to the number of COVID-19–related admissions. The feature preselectors F-score and mutual information were added later.

The predictions made by both models for data gathered between March 16, 2020, and June 15, 2020, were compared to the true admission count (Figure 1). The first 4 weeks were reserved for training the models before predictions were made in the week starting on April 16, 2020. For the consecutive weeks, the models were retrained with the data from the previous weeks. Both models made predictions with a similar error (linear regression: mean absolute error [MAE] 138, mean absolute percentage error [MAPE] 31; XGBoost regression: MAE 135, MAPE 31). The predictions become more accurate toward the end of the study period, since the models had more data to learn from, which reduced the sampling bias. May 22, 2020, was a public holiday in Finland, explaining a similar drop in the true admission count as on weekends.

To check that the Omaolo questionnaire data are relevant for the predictions, we compared the error of both models with and without the questionnaire data, and only with the questionnaire data (Table 2). Both models achieved the lowest error when the registry and the questionnaire data were combined, indicating that the questionnaire data are relevant for making accurate predictions compared to the registry data alone.

In addition to expert-selected features, we tested automated feature selection methods that select the top 8 features based on the F-score or mutual information of the feature with the predicted variable (number of admissions on a given day) (Table 3). The F-score and mutual information are measures of dependence between the feature values and the values of the predicted variable in the historical data. Different feature selection strategies worked better for the two models: linear regression was the most accurate with the mutual information criterion, whereas XGBoost was the most accurate with the expert-selected features.

Older age groups, who are more likely to have a severe form of COVID-19 and hence be admitted to hospital, are underrepresented in the Omaolo questionnaire data. This could affect the performance of the models, as it is more difficult for the models to learn from imbalanced training data. To assess the problem, oversampling of the underrepresented age groups was performed to see if it would decrease the error of the models (Table 4). Resampling gave the linear regression model slightly smaller error, whereas the XGBoost regressor performed worse with resampling. These results indicate that it is not essential when minimizing the prediction error to oversample the questionnaire answers of the underrepresented age groups to match the age distribution of the population.

Figure 1. COVID-19–related admissions predicted by linear regression and extreme gradient boosting (XGBoost) regression models, together with the true admission count during the first wave of the pandemic in 2020.
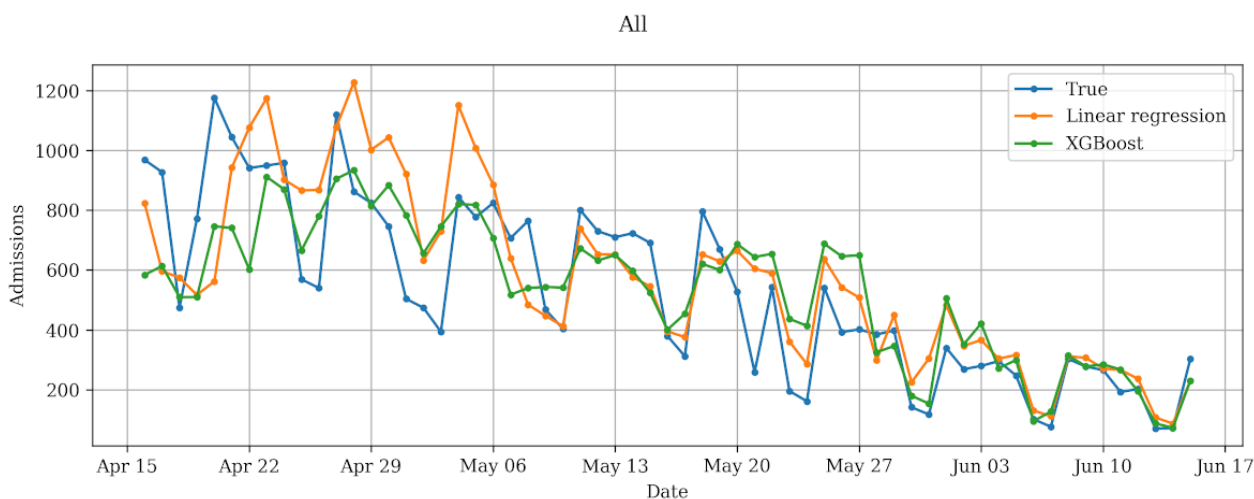
Table 2. Comparison of the effect of Omaolo[a] and Hilmo[b] data on the error of the models on expert-selected features.

Features	Mean absolute error	Mean absolute percentage error
Linear regression		
Omaolo+Hilmo	137.79	31.33
Omaolo	175.38	44.60
Hilmo	184.37	34.98
Extreme gradient boosting regression		
Omaolo+Hilmo	139.11	31.78
Omaolo	165.18	46.94
Hilmo	178.92	46.63

[a]Omaolo: A web-based, CE-marked symptom self-assessment tool and medical device.

[b]Hilmo: National administrative register on hospital admissions.

Table 3. Comparison of the effect of different feature selection methods on the error of the models.

Feature selection	Mean absolute error	Mean absolute percentage error
Linear regression		
Expert-selected	137.79	31.33
F-score	141.40	30.46
Mutual information	112.16	24.23
Extreme gradient boosting regression		
Expert-selected	139.11	31.78
F-score	150.07	36.40
Mutual information	146.12	33.33

Table 4. The effect of oversampling on the error of the models with expert-selected features.

Resampling	Mean absolute error	Mean absolute percentage error
Linear regression		
No resampling	137.79	31.33
Oversampling	131.13	30.47
Extreme gradient boosting regression		
No resampling	139.11	31.78
Oversampling	153.35	35.22

Profiling the Motives of the Omaolo COVID-19 Symptom Checker Users

The three different models produced the following sensitivity and specificity for detecting users that answered they were using the service out of curiosity: 0.622 and 0.367, respectively, for the naive Bayes classifier; 0.665 and 0.332, respectively, for logistic regression; and 0.607 and 0.388, respectively, for the XGBoost binary classifier. These results were acquired by maximizing the number of correct classifications.

Discussion

Principal Results

In this study, we examined whether it is possible to predict the national epidemic progression and the burden of health care using machine learning methods with real-life data on symptoms and usage of health care. The main finding of this study is that it is possible to predict national health care admissions related to COVID-19 using a symptom checker combined with register data by using machine learning methods with considerable accuracy (small MAPE error). These methods were tested in a scenario where the predictions were made 1 week ahead, once per week. The best model was achieved using the symptom checker data combined with register data (MAPE 24.23%). This result was reached by using linear regression with mutual

information as the feature preselector. All tested models and combinations of feature preselectors and models were able to produce predictions that followed the true epidemic progression (Figure 1). Overall, linear regression was better than XGBoost, although only marginally. This suggests that in our research scenario, there was no benefit in using a model that has many trainable hyperparameters (XGBoost regression) over a simple model (linear regression). All tested models seemed to improve toward the end of our study period as more data were available for training. Additionally, the differences in accuracy between the models were more visible at the start of the period and seemed to diminish toward the end. Based on the results, the F-score and especially mutual information appear to improve the results for linear regression. Feature preselection may improve the predictions by, for example, reducing the risk of overfitting. This is relevant in our data set since the feature set used in the classification was rather large and likely suffers from multicollinearity. Using the preselectors did not improve XGBoost regression. This suggests that we were not able to find a suitable preselection strategy for the method.

Finally, predictions that can follow the progression can be made using either Omaolo symptom checks or historic Hilmo counts separately. However, the best results were achieved by combining both. Adding Omaolo to Hilmo counts reduced the MAPE of linear regression from 34.98% to 31.33% and that of XGBoost from 46.63% to 31.78%. These results suggest that Omaolo contains information of the pandemic progression that is not present in Hilmo alone.

Oversampling the data to balance the regional differences between the Omaolo users and general population seemed to produce conflicting results: marginal gain with linear regression but loss of accuracy with XGBoost. Oversampling leads to added complexity in the analysis pipeline, and without a clear benefit, its use is hard to justify.

Profiling the Motives of the Omaolo COVID-19 Symptom Checker Users

The answer profiles of those using the COVID-19 symptom checker out of curiosity were very similar to those of the other users, and no reliable classification between the groups could be made by any of the tested models. Neither of the groups reported longer or more serious symptoms over the other. The only striking difference between the groups was that there appeared to be more out-of-curiosity responses during high service utilization such as after television infomercials.

Consideration of Other Sources

We also considered using the daily number of phone calls received at the 116117 Medical Helpline service. The 116117 Medical Helpline provides professional assistance on health care–related topics in urgent, but nonemergency cases to over 4 million Finns in extended business hours. However, since it was not possible to extract the topic of the call, whether they were COVID-19–related or not, and since the calls could be localized with much poorer resolution than with the rest of the sources, this data set was eventually dropped from the analyses.

The Google Trends [23] of popular COVID-19–related search terms was another potential data source considered for analyses.

However, it was found that the publicly available Finnish trends only covered major cities. Additionally, the overlap of trending search terms between cities was found to be small, making the data very sparse. For these reasons, this data set was not used in the analyses.

Strengths and Limitations

This study had several potential limitations. A considerable number of responses (132,951) were not saved during the study period. Nearly all of these were submitted during the first 2 weeks of the study, complicating the analysis for the first month. This may have contributed to the relatively poor prediction accuracy for the related weeks by delaying the convergence of the regressors to the true admission count.

The true admission count showed a strong diminishing trend toward the end of our study period. During the last weeks of the period, there were days when only a few dozen new admissions were recorded nationally. Because our error metric, MAPE, is relative to the true values, if these values are small, error values will appear high even though the absolute error between true and predicted counts remains low. Despite this, we decided to use MAPE for its intuitiveness, wide use, and easy comparability of the error between the days, weeks, and methods.

At the beginning, the survey did not include an item about the motive to fill in the survey (ie, whether it was due to actual symptoms or out of curiosity). This adds some additional forms to our data that do not reflect the situation at hand. The proportion of responses filled-in out of curiosity remained remarkably stable at around 40% throughout the study period. Moreover, the results of trying to separate the responses filled-in out of curiosity from the rest with binary classifiers (naive Bayes, logistic regression, XGBoost classifier) failed to reveal any meaningful differences between the answer profiles of these groups. Thus, we did not find a justification to remove these responses from the analyses or handle them differently than the rest of responses (not filled-in out of curiosity).

The data available for this study contained COVID-19–related admissions data with a steady downward trend, and it would have been interesting to see if the models could predict a reversal of the trend before it occurs. However, there is weekly variability in the admissions and the models learned this pattern well. An interesting deviation from the weekly pattern was Ascension Day on Thursday, May 21, which was a public holiday in Finland. The models did not have enough training data on admissions on public holidays on a weekday to predict a similar dip in the admissions as on weekends, although longer data sets could allow the models to learn this pattern as well.

Much of the prediction errors took place during days that showed sharp peaks of increased or decreased demand that were not immediately explainable with the data available. Some of these errors may be due to technical reasons such as a major care provider suffering an error on one day and reporting higher counts on the following day. Naturally, these kinds of special events cannot be learned from the admission count data alone. On a positive note, the developed models appear robust and thus not susceptible to these kinds of anomalies.

One could also question if biological tests make other monitoring redundant. While in many countries biological tests are performed to follow the pandemic, it is important to note that online symptom checking does not replace the need for biological testing but provides a different and valuable perspective. In many countries, including Finland, the testing services are saturated, and only some population groups get tested, making our picture of the pandemic progression biased. Biological testing further comes with an immense cost, particularly with an exponential rise in cases.

The symptom checker, in turn, can be filled in anytime and by large numbers. Furthermore, it can be continuously updated to include the most relevant questions. It is possible that the health care burden is going to change even quite rapidly, and we can see that using Omaolo surveys, even without an indication from the time-series data. Additionally, the machine learning models are trained every week, allowing adaptability to changes in the statistical relationship between the predictors and the predicted variable as the pandemic progresses.

Furthermore, while our data only cover the first wave of the pandemic, the results remain important from the perspective of early decision-making against a new threat, and overcoming the challenge on modeling a novel phenomenon from the start with no history. Additionally, we do not only validate the data, but our approach further enables the prediction and modeling of the pandemic. Emergence of vaccinations, new variants, new policies, and restrictions will all affect the progression of the pandemic. These changes have also affected Omaolo, which has been continuously updated (eg, including questions about vaccination status). Our prediction models are also upgradeable continuously; thus, the drift in data and concept can be mitigated on the go, and features that will drop or rise in importance can be monitored along with the actual result of the prediction.

Finally, the lack of reported anosmia could suggest a lack of specificity. In our data, anosmia is a rarely reported symptom. It is also a known problem that without a specific test, COVID-19 is notoriously hard to distinguish from a common flu. Despite the relatively small number of reported anosmia cases, we have other features that have been shown to be highly significant in predicting International Classification of Diseases-10– and International Classification of Primary Care-2–coded admissions. These include features such as how many of the questionnaires were filled-in by people over 60 years old, how many reported lengths of symptoms were greater than 10 days, and how many were assigned an urgency code P1 (the most urgent) in the care recommendations.

Alternatively, if anosmia is important for specificity, it is likely that we have more false positives and our final models are less accurate. In other words, with the key symptoms, the precision of the model is improved and the error is smaller. The consequence of such a lack depends on how the results are used. If the burden of health care were solely estimated based on these models, we might underestimate the need for health care. However, it is unlikely that the model would be used to define any absolute health care need, whereas it can provide an indication to prepare for an increased health care burden.

The study also has several strengths. We had access to a nationwide online symptom checker data source, Omaolo, which is a CE-marked medical device complying with the Medical Device Directive, used by health care and social service professionals [24]. We also had data on all hospital admission records from public or private hospitals in Finland on a weekly basis. Such data are rarely available anywhere in the world and provide unique opportunities to produce new information about the possibilities of using such real-life data in predicting a subsequent health care burden. Similar symptom checkers could be adopted for use in many other countries, and they could provide an opportunity to collect data on symptom development very rapidly and at a relatively low cost at a national level. These symptom checkers and the findings are not restricted or solely applicable to the current pandemic or its first wave, but could be applied in any other future epidemic or pandemic or for the collection of other types of symptoms as well.

In addition, a clear advantage of machine learning methods is that both the model section and fit are automatized. This means that the prediction method adapts in the face of any new data and attempts to make as accurate predictions as possible with the data in use at any given time. However, in case the phenomenon in question changes notably, changes in data use and sources and machine learning methods are also required.

Conclusions

Our study shows that COVID-19–related health care admissions in the short term can be predicted with considerable accuracy using symptom checker data combined with register data based on machine learning methods. This type of approach could help health care providers better assess the burden of the health care system in advance, which would make resource allocation more predictable. Furthermore, we consider that this type of approach could also be implemented in different stages of the pandemic and in future pandemics as well.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Symptom checker questionnaire.

[DOCX File , 17 KB - formative_v6i3e35181_app1.docx]

References

1. Tiirinki H, Tynkkynen L, Sovala M, Atkins S, Koivusalo M, Rautiainen P, et al. COVID-19 pandemic in Finland - Preliminary analysis on health system response and economic consequences. Health Policy Technol 2020 Dec;9(4):649-662 [FREE Full text] [doi: 10.1016/j.hlpt.2020.08.005] [Medline: 32874860]

2. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar 28;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]

3. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, China Novel Coronavirus Investigating Research Team. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]

4. Rolling updates on coronavirus disease (COVID-19). World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen [accessed 2021-10-24]

5. Jormanainen V, Kaikkonen R, Isosomppi S. The Finns quickly found the coronavirus symptom assessment online service: those who responded to the Omaolo coronavirus symptom assessment from 16 March to 15 May, 2020. Helsinki: Finnish Institute for Health and Welfare; 2020.

6. Judson TJ, Odisho AY, Neinstein AB, Chao J, Williams A, Miller C, et al. Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. J Am Med Inform Assoc 2020 Jun 01;27(6):860-866 [FREE Full text] [doi: 10.1093/jamia/ocaa051] [Medline: 32267928]

7. Galmiche S, Rahbe E, Fontanet A, Dinh A, Bénézit F, Lescure F, et al. Implementation of a self-triage web application for suspected COVID-19 and its impact on emergency call centers: observational study. J Med Internet Res 2020 Nov 23;22(11):e22924 [FREE Full text] [doi: 10.2196/22924] [Medline: 33147165]

8. Timmers T, Janssen L, Stohr J, Murk JL, Berrevoets MAH. Using eHealth to support COVID-19 education, self-assessment, and symptom monitoring in the Netherlands: observational study. JMIR Mhealth Uhealth 2020 Jun 23;8(6):e19822 [FREE Full text] [doi: 10.2196/19822] [Medline: 32516750]

9. Bastiani L, Fortunato L, Pieroni S, Bianchi F, Adorni F, Prinelli F, et al. Rapid COVID-19 screening based on self-reported symptoms: psychometric assessment and validation of the EPICOVID19 Short Diagnostic Scale. J Med Internet Res 2021 Jan 06;23(1):e23897 [FREE Full text] [doi: 10.2196/23897] [Medline: 33320825]

10. Rossman H, Keshet A, Shilo S, Gavrieli A, Bauman T, Cohen O, et al. A framework for identifying regional outbreak and spread of COVID-19 from one-minute population-wide surveys. Nat Med 2020 May;26(5):634-638 [FREE Full text] [doi: 10.1038/s41591-020-0857-9] [Medline: 32273611]

11. Drew DA, Nguyen LH, Steves CJ, Menni C, Freydin M, Varsavsky T, COPE Consortium. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. Science 2020 Jun 19;368(6497):1362-1367 [FREE Full text] [doi: 10.1126/science.abc0473] [Medline: 32371477]

12. Jormanainen V, Soininen L. Use and users of the web-based Omaolo Covid-19 symptom self-assesment tool in Finland since March 16, 2020. Stud Health Technol Inform 2021 May 27;281:739-743. [doi: 10.3233/SHTI210270] [Medline: 34042674]

13. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]

14. Sund R. Quality of the Finnish Hospital Discharge Register: a systematic review. Scand J Public Health 2012 Aug;40(6):505-515. [doi: 10.1177/1403494812456637] [Medline: 22899561]

15. Paavo database. Statistics Finland. URL: https://www.stat.fi/tup/paavo/index_en.html [accessed 2021-10-30]

16. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning with applications in R. New York: Springer; 2013.

17. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. 2016 Presented at: KDD '16 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco. [doi: 10.1145/2939672.2939785]

18. Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. Phys Rev E 2004 Jun 23;69(6):066138. [doi: 10.1103/physreve.69.066138]

19. Cochrane C. Time series nested cross validation. Towards Data Science. 2018 May 18. URL: https://towardsdatascience.com/time-series-nested-cross-validation-76adba623eb9 [accessed 2021-10-31]

20. McCallum A, Nigam K. A comparison of event models for Naive Bayes text classification. AAAI Technical Report WS-98-05. AAAI. URL: https://www.aaai.org/Papers/Workshops/1998/WS-98-05/WS98-05-007.pdf [accessed 2021-10-31]

21. Python. URL: https://www.python.org [accessed 2021-10-31]

22. Scikit-learn. URL: https://scikit-learn.org/stable/ [accessed 2021-10-31]

23. Google Trends. Google. URL: http://www.google.com/trends/ [accessed 2021-10-30]

24. Omaolo instructions for use. Omaolo. URL: https://www.omaolo.fi/kayttoohjeet/omaolo-instructions.pdf [accessed 2021-10-31]

XSL•FO

RenderX

Abbreviations

CE: Conformité Européenne
MAE: mean absolute error
MAPE: mean absolute percentage error
THL: Finnish Institute for Health and Welfare
XGBoost: extreme gradient boosting

XSL•FO
RenderX

Original Paper

Blood Pressure Control in Individuals With Hypertension Who Used a Digital, Personalized Nutrition Platform: Longitudinal Study

Shivani Bakre[1,2], BBA; Benjamin Shea[1,3], BS; Jason Langheier[1], MPH, MD; Emily A Hu[1], PhD

[1]Foodsmart, San Francisco, CA, United States

[2]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[3]Department of Biostatistics, Harvard T H Chan School of Public Health, Boston, MA, United States

Corresponding Author:
Emily A Hu, PhD
Foodsmart
595 Pacific Ave
Floor 4
San Francisco, CA, 94133
United States
Phone: 1 781 866 9200
Email: emily.hu@foodsmart.com

Abstract

Background: While there is a strong association between adhering to a healthy dietary pattern and reductions in blood pressure, adherence remains low. New technologies aimed to help facilitate behavior change may have an effect on reducing blood pressure among individuals with hypertension.

Objective: This study aims to evaluate characteristics of participants with stage 2 hypertension who used Foodsmart and to assess changes in systolic blood pressure (SBP) and diastolic blood pressure (DBP).

Methods: We analyzed demographic, dietary, and clinical characteristics collected from 11,934 adults with at least two blood pressure readings who used the Foodsmart platform. Stage 2 hypertension was defined as SBP ≥140 mmHg or DBP ≥90 mmHg. We calculated mean changes in blood pressure among participants with stage 2 hypertension and stratified by length of follow-up and the covariates associated with achieving blood pressure levels below stage 2 hypertension. We compared changes in diet quality and weight between participants with stage 2 hypertension at baseline who achieved stage 1 hypertension or below and those who did not.

Results: We found that 10.63% (1269/11,934) of participants had stage 2 hypertension at baseline. Among Foodsmart participants with stage 2 hypertension at baseline, SBP and DBP decreased, on average, by 5.7 and 4.0 mmHg, respectively; 33.02% (419/1269) of participants with stage 2 hypertension at baseline achieved blood pressure levels below stage 2 hypertension (SBP <140 mmHg and DBP <90 mmHg). Using a multivariable ordinal logistic regression model, changes in Nutriscore (*P*=.001) and weight (*P*=.04) were statistically significantly associated with changes in blood pressure categories for users with stage 2 hypertension at baseline. Using a multivariable logistic regression model, we found that baseline Nutriscore, change in Nutriscore, and change in weight were associated with greater likelihood of users with stage 2 hypertension at baseline achieving a lower blood pressure category.

Conclusions: This study evaluated changes in SBP and DBP among users (with hypertension) of the Foodsmart platform and found that those with stage 2 hypertension, on average, improved their blood pressure levels over time.

Introduction

Almost half of all adults living in the United States have hypertension, with only 1 in 4 having their condition under control [1]. Hypertension occurs when the body's blood vessels narrow, which results in blood exerting a greater pressure on the walls of one's blood vessels and the heart working harder [2]. Uncontrolled hypertension can also lead to a myriad of

XSL•FO

RenderX

health complications as it can result in narrow, weak, or thick blood vessels, which prevent organs from functioning normally. Notably, it can cause aneurysm, heart disease, or stroke, with heart disease and stroke being the leading causes of death in the United States [1,3]. Consequently, this condition costs the US health care system about US $131 billion annually [4].

A healthy diet is an important part of an individual's toolkit to reduce their blood pressure. Previous studies have shown that a healthy dietary pattern can reduce hypertension to controlled levels. For example, the Dietary Approaches to Stop Hypertension (DASH) diet is composed of fruits, vegetables, and low-fat dairy products. Filippou et al [5] found in a systematic review of studies that the DASH diet resulted in significant reductions in blood pressure both for those with hypertension and those without. After such success in these trials, DASH is now recommended by the National Institutes of Health (NIH) to reduce hypertension [6]. Despite robust evidence supporting the relationship between healthy eating and lower risk of hypertension, barriers related to time, finances, information, and accessibility in obtaining healthy foods prevent people from starting and sustaining such changes over their lifetime.

Foodsmart, a digital nutrition platform, is a potential solution to address these barriers. The platform helps its users improve their health by providing them information about their current diet and pinpointing areas of improvement, while also providing personalized recipes and meal planning options to achieve target goals. The platform also assists in purchasing healthier foods with direct integration from meal plans to grocery lists to retailers through an ad-free environment. Consequently, Foodsmart is able to create sustainable behavior change for its users by making the process easy, quick, and affordable. Previous research has shown that this sustained behavior change has resulted in sustained changes in biometrics as well. Hu et al [7] found that Foodsmart members with obesity have been able to achieve sustained weight loss over the time during which they have used the platform.

Previous studies have also shown the effectiveness of digital nutrition interventions in helping participants lower their blood pressure. Milani et al [8] provided a digital intervention that consisted of lifestyle recommendations as well as medication management to assist patients in reducing their blood pressure. Participants were asked to measure their blood pressure at least once a week on purchased blood pressure devices. They found that in the digital intervention groups patients saw an average decrease in blood pressure of 14 mmHg in systolic blood pressure (SBP) and 5 mmHg in diastolic blood pressure (DBP) after 90 days compared with those on usual care with a reduction of 4 mmHg/2 mmHg [8]. This study displays what is possible at the higher end of blood pressure reductions over 90 days given that this was a partially medication-guided intervention. Steinberg et al [9] implemented a digital intervention to attempt to increase DASH adherence and thus lower blood pressure more through the use of a mobile diet-tracking app. After 3 months, participants had an average reduction of 2.8 mmHg of SBP and 3.6 mmHg for DBP [9]. Foodsmart, however, does not provide medication management services and differs from the latter intervention in its digital interface, personalized meal

planning, and online food ordering system. Therefore, Foodsmart is well-positioned to help its participants with hypertension make sustainable, holistic changes in their health by both providing them with nutritional information and adjusting their food purchasing environment to facilitate behavior change.

This study aims to characterize Foodsmart participants with stage 2 hypertension, as well as identify what characteristics are associated with returning to stage 1 hypertension levels or lower among those with stage 2 hypertension. It also aims to evaluate changes in blood pressure levels and other biometrics over time among participants with hypertension.

Methods

Study Sample

As of August 2021, 106,816 participants of Foodsmart who were older than 18 years living the United States and enrolled since January 2016 had entered plausible systolic and diastolic readings (SBP between 80 and 300 mmHg, exclusive; DBP between 40 and 200 mmHg, exclusive). Of those Foodsmart participants, 11,934 had reported at least two systolic/diastolic readings, with the first and last reported readings at least 30 days apart. Our final sample size was 11,934 participants with at least two reports of SBP and DBP.

Foodsmart

Foodsmart is a digital nutrition platform that facilitates sustained dietary and behavior change through 2 main components, FoodSmart and FoodsMart. Foodsmart has been described previously [10]. In brief, participants on Foodsmart complete an online dietary assessment, Nutriquiz, to evaluate current diet quality. Based on the results of the Nutriquiz, personalized meal plans, aligned with their dietary preferences, are created that are converted into grocery lists and integrated into online order and delivery of meal kits, prepared foods, and groceries.

Foodsmart is available through health plans and employers and can be accessed via the web or the iOS or Android operating system.

Measurements of SBP, DBP, and Weight

Foodsmart participants were able to input their biometrics such as blood pressure, height, and weight, and were able to report new biometrics at any time. Because of potential errors in participants self-reporting their biometrics, the following values were considered as outliers and removed: SBP ≤80 mmHg or ≥300 mmHg, DBP ≤40 mmHg or ≥200 mmHg, weight ≤27.2 kg or ≥181.1 kg, and BMI ≤15 kg/m^2 or ≥50 kg/m^2. Length of follow-up was calculated as the number of days between the first and last systolic/diastolic measures inputted. We defined normotension as SBP <120 mmHg and DBP <80 mmHg, elevated blood pressure as SBP between 120 and 129 mmHg (inclusive) and DBP <80, stage 1 hypertension as either SBP between 130 and 139 mmHg (inclusive) or DBP between 80-89 mmHg (inclusive), and stage 2 hypertension as either SBP ≥140 mmHg or DBP ≥90 mmHg, as defined by the American Heart Association [11]. These rules were applied to both the first and last systolic/diastolic measures reported. We used the stage 2

XSL•FO

RenderX

hypertension cutoff for the primary analysis, and the stage 1 hypertension cutoff for the sensitivity analysis. Changes in blood pressure were calculated by subtracting the end SBP and DBP readings from the baseline SBP and DBP readings, respectively. Percent changes were calculated for both SBP and DBP readings by dividing the change in SBP/DBP readings by the baseline SBP/DBP measurements.

Baseline BMI was calculated as the baseline weight entered in kilograms divided by height in meters squared (kg/m^2). Participants' baseline BMI was categorized as normal BMI (BMI <25 kg/m^2), overweight (25-29.9 kg/m^2), or obese (BMI ≥30 kg/m^2). Participants were also able to report any conditions they currently had (eg, diabetes, high blood pressure) in the Nutriquiz.

Dietary Assessment

Participants self-reported their food intake and dietary habits in the Foodsmart platform. Participants were asked to fill out a 53-item food frequency questionnaire called Nutriquiz (adapted from the National Cancer Institute Diet History Questionnaire I) when they sign up [12]. Demographic information (age, sex, height), weight, and daily dietary intake (added sugars, fiber, fruits, vegetables, whole grains, fats, proteins, water, and sodium) were collected in the Nutriquiz.

Based on responses from the Nutriquiz and the daily dietary intake, a score (Nutriscore) from 0 to 70 was calculated to assess overall diet quality, with 70 being the highest diet quality. The Nutriscore is based on the Alternative Healthy Eating Index-2010 and the Commonwealth Scientific and Industrial Research Organization Healthy Diet Score [13,14]. Change in the Nutriscore was calculated as a participant's last Nutriscore minus the first Nutriscore. A positive change in Nutriscore indicates the participant improved their nutrition.

Statistical Analysis

We performed exploratory data analyses to assess the baseline demographic characteristics, SBP and DBP levels, and diet quality of the total study population, stratified by whether participants had stage 2 hypertension at baseline or not. We reported categorical variables as number of participants (percentage of study population) and continuous variables as mean (SD). We used chi-square tests to assess whether categorical variables are independent of baseline stage 2 hypertension status, and 2-sample t tests (unpaired) to evaluate differences in continuous variables.

Univariate and multivariable ordinal logistic regression analyses were used to estimate the odds ratios (ORs) and 95% CIs of achieving different blood pressure categories at different thresholds (eg, using the elevated blood pressure category as the threshold, the odds of achieving below stage 1 hypertension and stage 2 hypertension). Four blood pressure categories were considered: normotension, elevated blood pressure, stage 1 hypertension, and stage 2 hypertension. Univariate and multivariable logistic regression analyses were used to estimate the ORs and 95% CIs of achieving blood pressure levels below stage 2 hypertension. The multivariable logistic regression and multivariable ordinal logistic regression models were both mutually adjusted for sex, age category, baseline BMI category, baseline Nutriscore, change in Nutriscore, change in weight, baseline SBP, baseline DBP, diabetes, and high cholesterol.

Among participants who had stage 2 hypertension, we calculated the mean changes in SBP and DBP overall, and by time of follow-up at different thresholds (≥6, ≥12, and ≥24 months). We used paired t tests to evaluate whether the changes in SBP and DBP were statistically significant. In addition, we calculated the mean percentage change for SBP and DBP. In a sensitivity analysis, we used a threshold of stage 1 hypertension or higher (SBP ≥130 mmHg or DBP ≥80 mmHg) to calculate mean changes in HbA1c.

We also calculated the percentage of participants with stage 2 hypertension at baseline who achieved stage 1 hypertension or lower by the end of follow-up, and stratified by follow-up length.

To further explore the impact on SBP and DBP, we examined changes in weight, diet quality, SBP, and DBP, stratified by whether participants with stage 2 hypertension at baseline achieved stage 1 hypertension or lower (SBP <130 mmHg and DBP <80 mmHg) by the end of follow-up.

We considered P values less than .05 to be significant for all tests. R Studio version 1.4.1106 and R version 4.0.5 (R Foundation) were used for all analyses.

Ethical Approval

The study was declared exempt from institutional review board oversight by the Pearl Institutional Review Board given the retrospective design of the study and the less than minimal risk to participants.

Results

Participant Characteristics

Baseline demographics and biometrics of the total study sample stratified by a threshold of stage 2 hypertension status are shown in Table 1. Of participants with at least two blood pressure entries, 47.15% (5627/11,934) were classified as having stage 1 hypertension or higher at baseline, and 10.63% (1269/11,934) were classified as having stage 2 hypertension at baseline.

There were 11,934 participants included in the total study sample; 69.93% (8346/11,934) of the participants were female, 19.44% (2320/11,934) were 60 years or older, 68.46% (8170/11,934) were either overweight or obese, 17.80% (2124/11,934) reported having diabetes, and 37.91% (4254/11,934) reported having high cholesterol. The mean weight was 83.9 (SD 22.1) kg, the mean baseline Nutriscore was 33.8 (SD 8.5) points, and the mean change in the Nutriscore was 1.7 (SD 7.1) points. The mean follow-up length was 8.8 (SD 9.5) months and ranged from 1 to 65 months. Compared with participants who did not have stage 2 hypertension, participants who did have stage 2 hypertension were more likely to be male, to have a higher weight and BMI, to have a lower baseline Nutriscore, to have a longer length of follow-up, and to self-report having diabetes or high cholesterol. Participants with stage 2 hypertension at baseline were also more likely to have a higher increase in Nutriscore compared with participants

without stage 2 hypertension at baseline, although the difference was not statistically significant (*P*=.28).

To better understand what blood pressure category participants attained at the end of follow-up, we used ordinal logistic regression to examine the association of covariates with the end blood pressure categories among participants who had stage 2 hypertension at baseline (Table 2). Threshold 1 indicates the probability of stage 2 hypertension, stage 1 hypertension, or elevated blood pressure versus normotension. Threshold 2 indicates the probability of stage 2 hypertension or stage 1 hypertension versus elevated blood pressure or normotension. Threshold 3 indicates the probability of stage 2 hypertension versus stage 1 hypertension, elevated blood pressure, or normotension. In the univariate ordinal logistic regression

models, participants who were female were less likely to end in the elevated blood pressure or higher categories versus the normotensive blood pressure category on average (OR 0.75, 95% CI 0.59-0.96; *P*=.02). Participants with a higher baseline Nutriscore were less likely to end in the elevated blood pressure or higher categories versus the normotensive blood pressure category on average (OR 0.97, 95% CI 0.94-0.99; *P*=.01). Participants whose Nutriscore increased at the end of follow-up were less likely to end in the elevated blood pressure or higher categories versus the normotensive category (OR 0.96, 95% CI 0.93-0.99; *P*=.008). Participants who gained more weight at the end of follow-up were more likely to end in the elevated blood pressure or higher categories versus the normotensive category on average (OR 1.02, 95% CI 1.01-1.04; *P*=.004).

Table 1. Baseline characteristics of total study sample and by baseline hypertension status.

Characteristic	All participants (n=11,934)	Below stage 2 hypertension (n=10,665)	Stage 2 hypertension (n=1269)	*P* value[a]
Female, n (%)	8346 (69.93)	7539 (70.69)	807 (63.59)	<.001
Age (years), n (%)				<.001
<40	2929 (24.54)	2697 (25.29)	232 (18.28)	
40-59	6685 (56.02)	5927 (55.57)	758 (59.73)	
≥60s	2320 (19.44)	2041 (19.14)	279 (21.99)	
Weight (kg), n; mean (SD)	11,919; 83.9 (22.1)	10,656; 82.6 (21.6)	1263; 93.0 (24.2)	<.001
Change in weight (kg), n; mean (SD)	7934; –0.4 (7.5)	7111; –0.4 (7.3)	726; –0.7 (8.7)	<.001
BMI category, n (%)				<.001
Normal	3621 (30.34)	3418 (32.05)	203 (16.00)	
Overweight	3691 (30.93)	3339 (31.31)	352 (27.74)	
Obese	4479 (37.53)	3803 (35.66)	676 (53.27)	
Missing	143 (1.20)	105 (0.98)	38 (2.99)	
Baseline systolic blood pressure (mmHg), n; mean (SD)	11,934; 120 (13.0)	10,665; 118 (9.8)	1269; 140 (19.1)	<.001
Baseline diastolic blood pressure (mmHg), n; mean (SD)	11,934; 76.0 (9.6)	10,665; 74.4 (7.6)	1269; 89.5 (13.4)	<.001
Follow-up duration (months), n; mean (SD)	11,934; 8.8 (9.5)	10,665; 8.7 (9.3)	1269; 9.8 (10.9)	.03
Diabetes, n (%)	2124 (17.80)	1787 (16.76)	337 (26.56)	<.001
High cholesterol, n (%)	4254 (35.65)	3701 (34.70)	553 (43.58)	<.001
Baseline Nutriscore (0-70), n; mean (SD)	11,934; 33.8 (8.5)	10,665; 34.2 (8.5)	1269; 31.3 (8.5)	<.001
Change in Nutriscore, n; mean (SD)	11,551; 1.7 (7.1)	10,335; 1.6 (7.1)	1196; 1.9 (7.3)	.28

[a]Chi-square tests and 2-sample *t* tests (paired) were used to test differences for categorical and continuous variables, respectively.

XSL·FO

RenderX

Table 2. Association between predictors and likelihood of blood pressure categories at the end of follow-up in the univariate and multivariable ordinal logistic regression models.

Association	Univariate odds ratio (95% CI)	P value	Multivariable odds ratio (95% CI)	P value
Sex (female)	0.75 (0.59-0.96)	.02	0.82 (0.60-1.12)	.22
Age				
<40	1 (reference)		1 (reference)	
40-59	0.76 (0.55-1.05)	.09	0.79 (0.53-1.18)	.24
≥60	0.61 (0.42-0.88)	.009	0.68 (0.42-1.09)	.11
Baseline Nutriscore (0-70; per 2 points)	0.97 (0.94-0.99)	.01	0.96 (0.92-0.99)	.02
Change in Nutriscore (per 2 points)	0.96 (0.93-0.99)	.008	0.93 (0.89-0.97)	.001
Baseline BMI category				
Normal	1 (reference)		1 (reference)	
Overweight	0.96 (0.66-1.37)	.81	0.90 (0.56-1.45)	.67
Obese	0.93 (0.67-1.30)	.69	0.78 (0.50-1.23)	.29
Change in weight (kg)	1.02 (1.01-1.04)	.004	1.02 (1.00-1.04)	.04
Length of follow-up (per 6 months)	0.91 (0.86-0.97)	.002	0.97 (0.90-1.04)	.40
Diabetes	0.91 (0.71-1.18)	.50	1.14 (0.81-1.59)	.46
High cholesterol	0.84 (0.66-1.05)	.13	0.80 (0.59-1.08)	.15
Baseline systolic (per 5 mmHg)	0.98 (0.95-1.01)	.17	1.00 (0.96-1.05)	.96
Baseline diastolic (per 5 mmHg)	1.04 (1.00-1.09)	.07	1.05 (0.99-1.12)	.12

After adjusting for all other covariates in the multivariable ordinal logistic regression model, change in Nutriscore was inversely associated with ending with elevated blood pressure, stage 1 hypertension, or stage 2 hypertension (OR 0.93, 95% CI 0.89-0.97; *P*=.001); for every 2-point increase in Nutriscore, a participant was 7% less likely to end in the elevated blood pressure or higher categories versus the normotensive category on average. Change in weight was positively associated with ending in the elevated blood pressure or higher categories (OR 1.02, 95% CI 1.00-1.04; *P*=.04). For every 1-kg increase in weight, a participant was 2% more likely to end with elevated blood pressure or higher versus normotension, on average.

We further examined the association between baseline characteristics and odds of achieving a blood pressure category below stage 2 hypertension in the univariate and multivariable logistic regression models (Table 3). In the univariate regression models, participants who were female were 30% more likely to achieve an end blood pressure category below stage 2 hypertension than participants who were male (OR 1.30, 95% CI 1.02-1.67; *P*=.04). Participants classified in the age group 60 years and older were 68% more likely to achieve an end blood pressure category below stage 2 hypertension than participants classified in the less than 40-year-old group (OR 1.68, 95% CI 1.15-2.46; *P*=.007). Participants with a higher baseline Nutriscore were more likely to achieve an end blood pressure category below stage 2 hypertension (OR 1.04, 95% CI 1.01-1.06; *P*=.01). Participants with an increase in Nutriscore were more likely to achieve an end blood pressure category below stage 2 hypertension (OR 1.04, 95% CI 1.01-1.08; *P*=.01). Participants with a decrease in weight were more likely to achieve an end blood pressure category below stage 2 hypertension (OR 0.98, 95% CI 0.96-0.99; *P*=.003). Participants with a longer length of follow-up were more likely to achieve an end blood pressure category below stage 2 hypertension (OR 1.11, 95% CI 1.05-1.19; *P*=.001).

Table 3. Odds of achieving blood pressure below stage 2 hypertension for users with stage 2 hypertension at baseline in univariate and multivariable logistic regression analyses.

Variables	Univariate odds ratio (95% CI)	P value	Multivariable odds ratio (95%)	P value
Sex (female)	1.30 (1.02-1.67)	.04	1.16 (0.85-1.61)	.36
Age (years)				
<40	1 (reference)		1 (reference)	
40-59	1.36 (0.98-1.89)	.07	1.34 (0.89-2.03)	.17
≥60	1.68 (1.15-2.46)	.007	1.58 (0.96-2.61)	.07
Baseline Nutriscore (0-70; per 2 points)	1.04 (1.01-1.06)	.01	1.05 (1.01-1.10)	.01
Change in Nutriscore (per 2 points)	1.04 (1.01-1.08)	.01	1.07 (1.03-1.12)	.003
Baseline BMI category				
Normal	1 (reference)		1 (reference)	
Overweight	1.14 (0.79-1.67)	.48	1.21 (0.74-2.00)	.45
Obese	1.19 (0.85-1.68)	.31	1.43 (0.89-2.31)	.14
Change in weight (kg)	0.98 (0.96-0.99)	.003	0.98 (0.96-1.00)	.03
Length of follow-up (per 6 months)	1.11 (1.05-1.19)	.001	1.04 (0.97-1.13)	.26
Diabetes	1.15 (0.88-1.49)	.30	0.92 (0.64-1.30)	.62
High cholesterol	1.21 (0.96-1.54)	.11	1.23 (0.90-1.68)	.20
Baseline systolic (per 5 mmHg)	1.02 (0.99-1.06)	.14	1.01 (0.96-1.05)	.78
Baseline diastolic (per 5 mmHg)	0.96 (0.92-1.01)	.12	0.96 (0.90-1.02)	.23

After adjusting for all other covariates in the multivariable logistic regression model, we found that for every 2-point increase in baseline Nutriscore, a user would be 5% more likely to achieve an end blood pressure category below stage 2 hypertension (OR 1.05, 95% CI 1.01-1.10; P=.01). In addition, adjusting for all other covariates, for every 2-point increase in change in Nutriscore, a participant would be 7% more likely to achieve an end blood pressure category below stage 2 hypertension (OR 1.07, 95% CI 1.03-1.12; P=.003). Adjusting for all other covariates, for every 1-kg increase in change in weight, a participant would be 2% less likely to achieve an end blood pressure category below stage 2 hypertension (OR 0.98, 95% CI 0.96-1.00; P=.03).

Changes in SBP and DBP

Multimedia Appendices 1 and 2 show the mean and percent change in SBP and DBP, respectively, for participants who had stage 2 hypertension at baseline. Changes were calculated overall and by length of follow-up, at ≥6, ≥12, and ≥24 months. The mean changes in SBP for overall and subsetting to length of follow-up ≥6, ≥12, and ≥24 months were –5.7 (SD 14.6), –7.0 (SD 15.7), –8.0 (SD 14.3), and –7.4 (SD 13.3) mmHg, respectively. The mean percentage changes in SBP were –3.5% (SD 10.3%), –4.4% (SD 10.4%), –5.2% (SD 9.4%), and –4.8% (SD 8.6%) for overall, ≥6, ≥12, and ≥24 months, respectively. The mean changes in DBP for overall and subsetting to length of follow-up ≥6, ≥12, and ≥24 months were –4.0 (SD 10.6), –4.4 (SD 10.5), –5.0 (SD 9.8), and –4.3 (SD 10.1) mmHg, respectively. The mean percentage changes in DBP were –3.8% (SD 9.7%), –4.4% (SD 9.7%), –5.1% (SD 8.9%), and –4.3% (SD 8.9%) for overall, ≥6, ≥12, and ≥24 months, respectively. Using paired t tests, all changes between baseline and end SBP

and DBP readings were statistically significant at the P<.05 level. Systolic and diastolic BP for overall, ≥6, ≥12, and ≥24 months were all P<.001.

For participants who had stage 1 hypertension at baseline, the mean changes in SBP for overall and subsetting to length of follow-up ≥6, ≥12, and ≥24 months were –2.0 (SD 9.4), –2.4 (SD 10.1), –3.0 (SD 10.3), and –2.7 (SD 9.9) mmHg, respectively. Mean percentage changes in SBP were –1.3% (SD 6.9%), –1.6% (SD 7.2%), –2.0% (SD 7.4%), and –1.8% (SD 7.1%) for overall, ≥6, ≥12, and ≥24 months, respectively. The mean changes in DBP for overall and subsetting to length of follow-up ≥6, ≥12, and ≥24 months were –1.8 (SD 6.9), –2.1 (SD 6.9), –2.5 (SD 7.3), and –2.4 (SD 7.1) mmHg, respectively. The mean percentage changes in DBP were –1.9% (SD 7.4%), –2.3% (SD 7.5%), –2.7% (SD 8.1%), and –2.6% (SD 7.6%) for overall, ≥6, ≥12, and ≥24 months, respectively. Using paired t tests, all changes between baseline and end SBP and DBP readings were statistically significant at the P<.05 level. Systolic and diastolic BP for overall, ≥6, ≥12, and ≥24 months were all P<.001.

We evaluated the percentage of participants who had stage 2 hypertension at baseline who achieved blood pressure levels below stage 2 hypertension, systolic reading <140 mmHg and diastolic reading <90 mmHg, at the end of follow-up (Multimedia Appendix 3). Across all length-of follow-up periods, 33.02% (419/1269) of participants with stage 2 hypertension at baseline lowered their blood pressure to below stage 2 hypertension by the end of follow-up. Among participants whose follow-up time was longer than 6, 12, and 24 months, 37.6% (238/633), 42.5% (131/308), and 38.4%

(66/172) of participants achieved blood pressure levels below stage 2 hypertension by the end of follow-up.

We examined the change in weight and Nutriscore for participants who lowered their blood pressure levels to below stage 2 hypertension versus those who remained in the stage 2 hypertension category (Multimedia Appendix 4). Participants who changed from stage 2 hypertension to below stage 2 hypertension lost 1.8 more kilograms in weight (0.0 kg versus 1.8 kg lost) and had 1.2 points higher (1.5-point increase versus 2.7-point increase) in their diet quality (Nutriscore) than those who remained in the stage 2 hypertension category.

Discussion

In this study, we found that 10.63% (1269/11,934) of Foodsmart participants had stage 2 hypertension at baseline. Foodsmart participants with stage 2 hypertension at baseline were more likely to be male, older, have a higher weight and BMI, have additional comorbidities (diabetes, high cholesterol), have a lower Nutriscore, and have a longer follow-up. For participants with stage 2 hypertension, blood pressure reduced by 5.7 mmHg/4.0 mmHg on average over a follow-up of 9.8 (SD 10.9) months. Among participants with stage 2 hypertension at baseline, 33.02% (419/1269) lowered their blood pressure to stage 1 hypertension or below. Furthermore, there were greater reductions in weight changes and increases in Nutriscores among those with stage 2 hypertension at baseline who achieved blood pressure levels below stage 2 hypertension. Our findings suggest that participant use of Foodsmart may be associated with reductions in blood pressure for patients with stage 2 hypertension.

Previous research confirms several of our findings. Through Korea's National Health and Nutrition Examination Survey, Noh et al [15] confirmed that, in line with our baseline analysis, those with hypertension are more likely to have additional comorbidities, noting that these comorbidities (obesity, dyslipidemia, and impaired fasting glucose) are also likely to be associated with age and male sex. Additional studies have provided further context to these findings. Weight, in particular, has been long-established to increase the risk of hypertension [16]. Having a BMI greater than or equal to 30 kg/m^2 (categorized as obese) has been associated with a 3.5 times higher likelihood of developing hypertension. In addition, previous research has shown that diabetes and hypertension often occur together because they are mediated by similar physiologic conditions, such as impaired circadian blood pressure rhythm, renin-angiotensin-aldosterone blood pressure system dysregulation, microvascular and macrovascular damage, and more [17,18]. Finally, prior research also suggests that dyslipidemia may be a risk factor for hypertension [19]. Dyslipidemia is known to cause endothelial damage, which results in the constriction of cardiovascular vessels as opposed to their dilation. This damage may consequently result in hypertension. In line with our findings, Neter et al [20] found a dose-response relationship between weight change and blood pressure reduction. They found a reduction in blood pressure of −1.05 mmHg/−0.92 mmHg per kilogram of weight loss. In addition, a meta-analysis on the DASH diet found that improved

nutrition is associated with greater decreases in blood pressure, with the mean difference of SBP and DBP being −3.2 and −2.5 mmHg, respectively, from a systematic review of 30 randomized clinical trials [5].

Individuals with hypertension are estimated to incur US $1920 higher incremental health care expenditures and nearly triple the medication expenditures compared with individuals without hypertension [4]. Drugs such as angiotensin-converting enzyme (ACE) inhibitors and calcium channel blockers account for a large proportion of these expenditures. Given the high cost of medications, prevention and management of hypertension through healthier eating could supplement hypertension drugs and potentially lower its usage. Unfortunately, we do not know whether participants were on blood pressure medications before or during enrollment on the Foodsmart platform. Despite this, we can estimate the difference in costs between prescription medications and Foodsmart. As of 2021, the Foodsmart platform on average costs US $12.30 per eligible member annually. Using the results above, a 5-mmHg reduction in SBP and DBP would cost US $12.30 on average for participants on the Foodsmart platform for at least a year. Using lisinopril (ACE inhibitor), by contrast, would cost US $48 on average to reduce SBP/DBP by 21.4/15.7 mmHg [21]. The Foodsmart platform could be a cost-effective program to complement the use of standard hypertension medications.

There are some important limitations to note for this study. The first is that blood pressure values were self-reported and not clinically validated. Given that only blood pressure values were reported, we do not have information on which blood pressure devices were used or if standard protocol, such as the average of 3 consecutive measurements, was used. We assumed participants inputted their last recorded blood pressure values from a clinical setting. Nevertheless, we have reason to believe these values should still be approximately accurate, because blood pressure is not a required input for the app, and therefore those who did input blood pressure were most likely motivated by personal tracking. In addition, because values were self-reported, follow-up time was based on when the biometrics were entered as opposed to when blood pressure was measured. Therefore, the first measurements of SBP and DBP recorded may not necessarily align with the beginning of Foodsmart enrollment. Another possible issue is selection bias for individuals with hypertension who use Foodsmart. Those with hypertension who use the app may be more likely to make lifestyle changes because they are using the app for the purposes of improving their nutrition. These users may also be working to make lifestyle changes outside of the app that affect their blood pressure. Therefore, we cannot state that Foodsmart caused these changes in blood pressure. To make statements about causation, a randomized controlled trial must be conducted. Next, we were not able to collect data on other factors that might affect hypertension status. For instance, participants' personal or family medical histories may have an influence on hypertension status. We lack information on whether participants were taking blood pressure medications before or during usage of the Foodsmart platform or if participants changed blood pressure medication during enrollment. Finally, we did not have data on socioeconomic

factors, such as education, which may confound the associations. Further research is necessary to understand these additional variables. We also did not examine the frequency of Foodsmart app use, which may affect the associations found in our analyses.

Our analysis also has many strengths. This is the first study to examine the real-life impact of behavioral change as a result of online food ordering, dietary education, and meal planning through a digital intervention and its impact on hypertension. Through analyzing data from Foodsmart's large user base, we were able to gather information about real-world associations between changes in diet and blood pressure levels. Furthermore,

Foodsmart participants had a wide range of enrollment lengths, which allowed us to evaluate changes in blood pressure over different lengths of time, including follow-up time of more than 2 years.

In summary, this study assessed changes in self-reported SBP and DBP for participants using a digital nutrition platform that offers personalized meal planning, food ordering, grocery discounts, and price comparisons. Further research through a randomized clinical trial is warranted to evaluate the causal effect between the use of the Foodsmart platform and changes in blood pressure among participants.

Authors' Contributions

BS acquired data, analyzed the data, interpreted the results, and drafted the manuscript. EAH designed the study, interpreted results, and drafted the manuscript. SB interpreted results and drafted the manuscript. JL interpreted results. All the authors reviewed and approved the final version of the manuscript and take responsibility for the manuscript.

Conflicts of Interest

SB, BS, JL, and EAH are employees of Foodsmart.

Multimedia Appendix 1
Mean change in SBP among participants who had stage 2 hypertension at baseline.
[PNG File , 72 KB - formative_v6i3e35503_app1.png]

Multimedia Appendix 2
Mean change in DBP among participants who had stage 2 hypertension at baseline.
[PNG File , 67 KB - formative_v6i3e35503_app2.png]

Multimedia Appendix 3
Percent of users who changed from stage 2 hypertension at baseline to below stage 2 hypertension at the end of follow-up.
[PNG File , 54 KB - formative_v6i3e35503_app3.png]

Multimedia Appendix 4
Change in biometrics stratified by whether participants with stage 2 hypertension at baseline achieved a blood pressure below stage 2 hypertension.
[PNG File , 41 KB - formative_v6i3e35503_app4.png]

References

1. Centers for Disease Control and Prevention. Facts about Hypertension. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/bloodpressure/facts.htm [accessed 2021-07-19]
2. Winters W. Hypertension. Encyclopædia Britannica. URL: https://www.britannica.com/science/hypertension [accessed 2021-07-21]
3. Mayo Clinic. High Blood Pressure. Mayo Clinic. URL: https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/symptoms-causes/syc-20373410 [accessed 2021-07-01]
4. Kirkland EB, Heincelman M, Bishu KG, Schumann SO, Schreiner A, Axon RN, et al. Trends in Healthcare Expenditures Among US Adults With Hypertension: National Estimates, 2003-2014. J Am Heart Assoc 2018 May 30;7(11):e008731 [FREE Full text] [doi: 10.1161/JAHA.118.008731] [Medline: 29848493]
5. Filippou CD, Tsioufis CP, Thomopoulos CG, Mihas CC, Dimitriadis KS, Sotiropoulou LI, et al. Dietary Approaches to Stop Hypertension (DASH) Diet and Blood Pressure Reduction in Adults with and without Hypertension: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. Adv Nutr 2020 Sep 01;11(5):1150-1160 [FREE Full text] [doi: 10.1093/advances/nmaa041] [Medline: 32330233]

XSL•FO
RenderX

6.   National Institutes of Health National Heart, Lung, and Blood Institute. Your Guide to Lowering Your Blood Pressure With DASH. National Institutes of Health National Heart, Lung, and Blood Institute. 2006. URL: https://www.nhlbi.nih.gov/files/docs/public/heart/new_dash.pdf [accessed 2021-07-01]

7.   Hu EA, Pasupuleti M, Nguyen V, Langheier J, Shurney D. Sustaining weight loss among adults with obesity using a digital meal planning and food purchasing platform for 12, 24, and 36 months: a longitudinal study. Nutr J 2021 Jan 21;20(1):8 [FREE Full text] [doi: 10.1186/s12937-021-00666-9] [Medline: 33478516]

8.   Milani RV, Lavie CJ, Bober RM, Milani AR, Ventura HO. Improving Hypertension Control and Patient Engagement Using Digital Tools. Am J Med 2017 Jan;130(1):14-20. [doi: 10.1016/j.amjmed.2016.07.029] [Medline: 27591179]

9.   Steinberg DM, Kay MC, Svetkey LP, Askew S, Christy J, Burroughs J, et al. Feasibility of a Digital Health Intervention to Improve Diet Quality Among Women With High Blood Pressure: Randomized Controlled Feasibility Trial. JMIR Mhealth Uhealth 2020 Dec 07;8(12):e17536 [FREE Full text] [doi: 10.2196/17536] [Medline: 33284116]

10.   Hu EA, Nguyen V, Langheier J, Shurney D. Weight Reduction Through a Digital Nutrition and Food Purchasing Platform Among Users With Obesity: Longitudinal Study. J Med Internet Res 2020 Sep 02;22(9):e19634 [FREE Full text] [doi: 10.2196/19634] [Medline: 32792332]

11.   Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension 2018 Jun;71(6):1269-1324 [FREE Full text] [doi: 10.1161/HYP.0000000000000066] [Medline: 29133354]

12.   Subar A, Thompson F, Kipnis V, Midthune D, Hurwitz P, McNutt S, et al. Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires : the Eating at America's Table Study. Am J Epidemiol 2001 Dec 15;154(12):1089-1099. [doi: 10.1093/aje/154.12.1089] [Medline: 11744511]

13.   Chiuve S, Fung T, Rimm E, Hu F, McCullough M, Wang M, et al. Alternative dietary indices both strongly predict risk of chronic disease. J Nutr 2012 Jun;142(6):1009-1018 [FREE Full text] [doi: 10.3945/jn.111.157222] [Medline: 22513989]

14.   Hendrie G, Baird D, Golley R, Noakes M. The CSIRO Healthy Diet Score: An Online Survey to Estimate Compliance with the Australian Dietary Guidelines. Nutrients 2017 Jan 09;9(1):47 [FREE Full text] [doi: 10.3390/nu9010047] [Medline: 28075355]

15.   Noh J, Kim HC, Shin A, Yeom H, Jang S, Lee JH, et al. Prevalence of Comorbidity among People with Hypertension: The Korea National Health and Nutrition Examination Survey 2007-2013. Korean Circ J 2016 Sep;46(5):672-680 [FREE Full text] [doi: 10.4070/kcj.2016.46.5.672] [Medline: 27721859]

16.   Mokdad AH, Ford ES, Bowman BA, Dietz WH, Vinicor F, Bales VS, et al. Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. JAMA 2003 Jan 01;289(1):76-79. [doi: 10.1001/jama.289.1.76] [Medline: 12503980]

17.   Schutta MH. Diabetes and hypertension: epidemiology of the relationship and pathophysiology of factors associated with these comorbid conditions. J Cardiometab Syndr 2007;2(2):124-130. [doi: 10.1111/j.1559-4564.2007.06368.x] [Medline: 17684469]

18.   Sowers JR. Insulin resistance and hypertension. Am J Physiol Heart Circ Physiol 2004 May;286(5):H1597-H1602 [FREE Full text] [doi: 10.1152/ajpheart.00026.2004] [Medline: 15072967]

19.   Halperin RO, Sesso HD, Ma J, Buring JE, Stampfer MJ, Gaziano JM. Dyslipidemia and the risk of incident hypertension in men. Hypertension 2006 Jan;47(1):45-50. [doi: 10.1161/01.HYP.0000196306.42418.0e] [Medline: 16344375]

20.   Neter JE, Stam BE, Kok FJ, Grobbee DE, Geleijnse JM. Influence of weight reduction on blood pressure: a meta-analysis of randomized controlled trials. Hypertension 2003 Nov;42(5):878-884. [doi: 10.1161/01.HYP.0000094221.86888.AE] [Medline: 12975389]

21.   Fogari R, Mugellini A, Zoppi A, Lazzari P, Destro M, Rinaldi A, et al. Effect of successful hypertension control by manidipine or lisinopril on albuminuria and left ventricular mass in diabetic hypertensive patients with microalbuminuria. Eur J Clin Pharmacol 2005 Aug;61(7):483-490. [doi: 10.1007/s00228-005-0961-2] [Medline: 16021438]

Abbreviations

ACE: angiotensin-converting enzyme
DASH: Dietary Approaches to Stop Hypertension
DBP: diastolic blood pressure
NIH: National Institutes of Health
OR: odds ratio
SBP: systolic blood pressure

XSL•FO
RenderX

XSL·FO
RenderX

Original Paper

Predicting Acceptance of e–Mental Health Interventions in Patients With Obesity by Using an Extended Unified Theory of Acceptance Model: Cross-sectional Study

Vanessa Rentrop[1], MSc; Mirjam Damerau[1], MSc; Adam Schweda[1], PhD; Jasmin Steinbach[1], MSc; Lynik Chantal Schüren[2]; Marco Niedergethmann[2], Prof Dr; Eva-Maria Skoda[1], PD, MD; Martin Teufel[1], Prof Dr; Alexander Bäuerle[1], PhD

[1]Clinic for Psychosomatic Medicine and Psychotherapy, LVR-University Hospital Essen, University of Duisburg-Essen, Essen, Germany
[2]Department of General and Visceral Surgery, Alfried-Krupp Hospital Essen, Essen, Germany

Corresponding Author:
Vanessa Rentrop, MSc
Clinic for Psychosomatic Medicine and Psychotherapy
LVR-University Hospital Essen
University of Duisburg-Essen
Virchowstraße 174
Essen, 45147
Germany
Phone: 49 201438755101
Email: vanessa.rentrop@uni-due.de

Abstract

Background: The rapid increase in the number of people who are overweight and obese is a worldwide health problem. Obesity is often associated with physiological and mental health burdens. Owing to several barriers to face-to-face psychotherapy, a promising approach is to exploit recent developments and implement innovative e–mental health interventions that offer various benefits to patients with obesity and to the health care system.

Objective: This study aims to assess the acceptance of e–mental health interventions in patients with obesity and explore its influencing predictors. In addition, the well-established Unified Theory of Acceptance and Use of Technology (UTAUT) model is compared with an extended UTAUT model in terms of variance explanation of acceptance.

Methods: A cross-sectional web-based survey study was conducted from July 2020 to January 2021 in Germany. Eligibility requirements were adult age (≥18 years), internet access, good command of the German language, and BMI >30 kg/m^2 (obesity). A total of 448 patients with obesity (grades I, II, and III) were recruited via specialized social media platforms. The impact of various sociodemographic, medical, and mental health characteristics was assessed. eHealth-related data and acceptance of e–mental health interventions were examined using a modified questionnaire based on the UTAUT.

Results: Overall, the acceptance of e–mental health interventions in patients with obesity was moderate (mean 3.18, SD 1.11). Significant differences in the acceptance of e–mental health interventions among patients with obesity exist, depending on the grade of obesity, age, sex, occupational status, and mental health status. In an extended UTAUT regression model, acceptance was significantly predicted by the *depression score* (*Patient Health Questionnaire-8*; β=.07; *P*=.03), *stress owing to constant availability via mobile phone or email* (β=.06; *P*=.02), and *confidence in using digital media* (β=−0.058; *P*=.04) and by the UTAUT core predictors *performance expectancy* (β=.45; *P*<.001), *effort expectancy* (β=.22; *P*<.001), and *social influence* (β=.27; *P*<.001). The comparison between an extended UTAUT model (16 predictors) and the restrictive UTAUT model (*performance expectancy*, *effort expectancy*, and *social influence*) revealed a significant difference in explained variance ($F_{13,431}$=2.366; *P*=.005).

Conclusions: The UTAUT model has proven to be a valuable instrument to predict the acceptance of e–mental health interventions in patients with obesity. The extended UTAUT model explained a significantly high percentage of variance in acceptance (in total 73.6%). On the basis of the strong association between acceptance and future use, new interventions should focus on these UTAUT predictors to promote the establishment of effective e–mental health interventions for patients with obesity who experience mental health burdens.

Introduction

Background

The prevalence of people who are overweight and obese has approximately tripled since 1975, resulting in >1.9 billion adults worldwide being obese in 2016 [1]. Taking this figure and the association of obesity with serious health complications into account, obesity is considered to be a global public health crisis [2]. Obesity increases the risk of noncommunicable diseases, such as cardiovascular disease, type 2 diabetes, and even some types of cancer [3,4] and increases the chance of remaining dependent on care in nursing homes [5]. Besides several physical comorbidities, obesity is also associated with psychological distress. It elevates the risk of mental health disorders, including depression, adjustment disorders, and anxiety disorders [6-8]. For instance, previous research found that more than half of the patients with obesity experienced at least one mental disorder, resulting in low quality of life and reduced self-esteem [9,10]. Furthermore, it was shown that a high number of people with obesity have an increased risk of developing depressive disorders [11] and show symptoms of major depression [12]. In particular, discrimination against patients with obesity and weight-related stigmatization play a central role and mediate the connection to negative mental and physical health outcomes and often lead to low quality of life and poor well-being [13]. It is important to note that distress and depression inversely increase the incidence of obesity again. A very central reason for this is that individuals with obesity are often blamed for their weight and a perception among the general population is that weight stigma is justified and can motivate individuals to adopt healthy behaviors [13]. Indeed, and speaking in terms of money, obesity causes large economic burdens for health care systems [14].

However, effective interventions for obesity management are scarce because they are often impractical to implement in the health care system [15]. The social stigma that accompanies obesity and the impediments in mobility of older patients with obesity are only 2 of many barriers preventing good professional patient care [15]. In particular, long-term psychotherapeutic and psychosocial care of people with obesity are often not possible owing to limited health care opportunities. As a result, the development and implementation of cost-effective and low-threshold approaches to handle elevated psychological distress and existing mental health disorders are essential.

E–mental health interventions are an effective and innovative approach that can circumvent the aforementioned barriers hindering the health care system. Especially during the COVID-19 pandemic, the need for contact-free, low-threshold, and easily accessible approaches in health care has become increasingly evident [16]. During this crisis, many institutions in the health care system tried to develop possibilities to serve their patients via videoconferencing or other digital methods [17,18]. Innovative e–mental health interventions offer a low-threshold, time- and location-flexible, and often anonymous alternative to traditional face-to-face therapy. Although the implementation of e–mental health interventions in Germany is in its early stage [19], multiple studies have shown effects comparable with that of face-to-face therapy [19,20]. Existing e–mental health interventions for patients with a variety of mental disorders mostly include psychoeducational interventions and web-based tasks designed for cognitive behavioral, psychodynamic, or acceptance and commitment therapy [21]. Acceptance of participants who participated in such interventions and their satisfaction with such interventions has often been considerable [21,22]. A recent study revealed that patients who had undergone bariatric surgery were very positive about eHealth interventions in their follow-up care [23]. Particularly after bariatric surgery, eHealth interventions can be effective in postoperative weight maintenance and in the reduction of eating disorder symptoms [24]. Although the proven user acceptance for e–mental health interventions [25] and the reasonable assumption that patients with obesity affected by mental health disorders would benefit from easily accessible and effective e–mental health interventions to support their mental well-being, the current opportunities are still in their infancy. Therefore, it is important to explore the patient-specific needs because newly developed interventions need to be tailored to these needs to foster user acceptance and treatment adherence [26].

Besides the advantages and resources described above, there are barriers to using eHealth interventions that should not be neglected, which will now be examined in more detail. In addition, previous research has shown that various predictors seem to influence patient acceptance of e–mental health interventions. A cross-sectional survey in 2016 demonstrated that the acceptance of web-based aftercare among inpatients (groups mentioned below) is low [21]. The highlighted factors that were able to significantly predict acceptance were social influence (SI), performance expectancy (PE), and effort expectancy (EE; 3 predictors of the Unified Theory of Acceptance and Use of Technology [UTAUT] model, which will be described in more detail in the following sections). Although high acceptance correlated with (young) age, (high) education, (high) level of information, and experience, the stress caused by permanent availability was associated with low acceptance; however, the effect was very small. Other factors influencing the acceptance of web-based aftercare were different patient groups (psychosomatic, cardiologic, orthopedic, pediatric, and substance-related disorders) [21] and current employment status [27].

In addition, 2 research studies focusing on the patient's perspective explored the following barriers of e–mental health interventions reported by health care providers: the lack of guidance through therapeutic relationship, limitation of communication, and control or concerns about data security [28,29]. To minimize the supposed disadvantages of e–mental

health interventions for patients and health care professionals and to focus on possible advantages, specific research in this area is necessary. As the implementation and uptake of e–mental health interventions are still low, the acceptance of such interventions and the barriers hindering them make their use less likely and need to be further evaluated [30]. Most studies assessing the acceptance of eHealth interventions are not based on valid measurement constructs [31,32]; therefore, for this study, it is important to determine that the results are evaluated based on solid operationalization and theoretical ramifications. Therefore, the UTAUT is used in this study [33], as it has already been used in research on eHealth to validly identify the determinants of acceptance [33]. The UTAUT model consists of four main predictors: *PE, EE, SI, and facilitating conditions (FCs)* [34]. It can be used to assess acceptance of an eHealth intervention or other technological systems using the first 3 core predictors, whereas acceptance itself is operationalized as *behavioral intention (BI)* to use such interventions on a fifth scale. PE describes the degree to which an individual believes that he or she will benefit from using the intervention. EE is defined as the degree of ease associated with the use of the technology. A person's assessment of the extent to which their relevant social contacts (ie, family and friends) would approve the intervention is indicated by SI. FC is defined as the degree to which an individual believes that an organizational and technical infrastructure exists to support the use of the system [33]. The acceptance (intention to use; Figure 1) is predicted by the three main predictors (*PE, EE, and SI*), whereas actual use behavior is predicted by the intention to use and FCs.

Researchers have pointed out that the UTAUT model needs to be explored and validated in different target groups [21]. A study using the UTAUT model for acceptance measurement conducted by Hennemann, Beutel, and Zwerenz in 2017 [30] revealed that acceptance of e–mental health interventions was quite low and acceptance of web-based aftercare was moderate and did not differ between age groups. Significant predictors of high acceptance are PE, SI, and treatment-related internet and mobile use [30]. A further study in 2019 found that FCs and perceived usefulness were associated with increased eHealth activity, whereas SI was not associated with eHealth use [35]. In 2015, de Veer et al [36] found that 63.1% of elderly people (aged between 57 and 77 years) included in their study would use an eHealth app (moderate acceptance) by using the UTAUT model. In this case, the model showed that PE, EE, and self-efficacy were highly related to acceptance of eHealth intervention, whereas SI was not [36]. Although previous research has focused on the acceptance of e–mental health interventions, it has rarely measured acceptance using validated constructs, making it important to use the UTAUT model in future research. In addition, because of the various findings to date on the different variables that might predict acceptance, more research with the UTAUT model is necessary. Research studies are needed, on the one hand, to establish relationships between the individual components, the use behavior, and the acceptance of e–mental health interventions in patients with obesity and, on the other hand, to validate the UTAUT model in other patient cohorts than the ones studied so far.

Figure 1. Unified Theory of Acceptance and Use of Technology (UTAUT) model and the extended model predictors including hypotheses (H) 1-5.
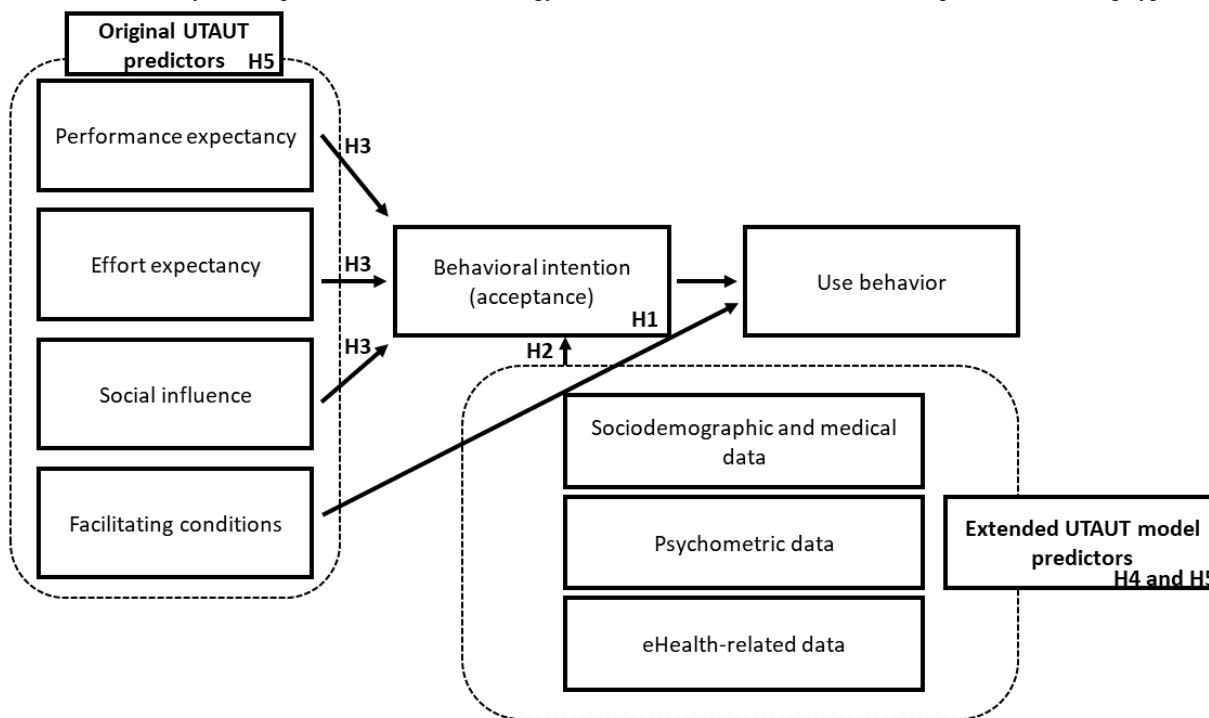
Objectives

This study aims to assess the acceptance of e–mental health interventions in patients with obesity (grades I, II, and III) and to explore the underlying, influencing factors determining the

acceptance. We use the established UTAUT model and extend it to accomplish the abovementioned goal. Previous studies have already shown that acceptance is associated with sociodemographic variables such as age and sex [21,37]. In addition, there were differences in acceptance depending on the

mental health status of the patients [37,38] and evidence that patients who have already undergone bariatric surgery show high acceptance of eHealth interventions [21]. Previous research has not yet addressed patients with obesity and their acceptance of e–mental health interventions. To identify additional variables associated with acceptance of e–mental health interventions, this study includes several obesity-specific factors (eg, bariatric surgery and the grade of obesity). Previous research has examined acceptance and various predictors in other patient groups, leading us to propose the following assumptions for our study:

- Hypothesis 1: In accordance with previous research in other patient groups [21,30,38], it is assumed that the overall acceptance of e–mental health interventions in patients with obesity is moderate.
- Hypothesis 2: Moreover, we assume that we will find group differences in the acceptance of e–mental health interventions depending on sex, age, grade of obesity, occupational status, mental disorder, outpatient psychotherapy, and previously performed bariatric surgery [21,25,27].
- Hypothesis 3: We postulate a positive relation between the UTAUT factors (SI, PE, and EE) and the acceptance of e–mental health interventions for people with obesity [21,33,35,36] (Figure 1).
- Hypothesis 4: Furthermore, we assume that, in addition to sociodemographic and medical factors, psychometric data and eHealth-related data, for example, internet anxiety (negative) and experience with e–mental health interventions (positive), significantly explain variance in the acceptance of e–mental health interventions among patients with obesity [21].
- Hypothesis 5: In a comparison between the restrictive UTAUT model and our extended UTAUT model, we hypothesize a significant difference in the explanation of variance.

The results of this study could significantly accelerate the process of implementing and adapting e–mental health interventions for specific patient groups, such as patients with obesity. Especially considering how the current pandemic has caused additional mental health burdens, more efficient and easily available ways to provide psychological support should be developed.

Methods

Study Design and Participants

A cross-sectional approach was implemented to measure the acceptance of e–mental health interventions and its underlying predictors in a sample of patients with obesity based on the UTAUT.

Participants were recruited from July 2020 to January 2021 at the Obesity Center of Alfried Krupp Hospital and via social media platform groups such as Facebook, exclusively directed toward patients who are seeking, undergoing, or have already undergone bariatric surgery and were aged ≥18 years. Other eligibility requirements were good command of the German language, internet access, and BMI >30 kg/m^2 (diagnosis of obesity). The classification of obesity grades according to the World Health Organization is as follows: (1) obesity grade I (BMI 30-34.9 kg/m^2), (2) obesity grade II (BMI 35-39.9 kg/m^2), and (3) obesity grade III (BMI ≥40 kg/m^2) [39]. The processing time of the web-based survey, consisting of 68 items in total, was approximately 18 minutes. No financial compensation was offered. Electronic informed consent was obtained before the survey began, and participation was completely anonymous and voluntary. Of 996 participants who started the survey, 643 (64.6%) participants completed it. A total of 30.3% (195/643) of participants were underweight, normal weight, or overweight but not with BMI >30 kg/m^2; thus, they were excluded from this study. This resulted in a total sample of 69.7% (448/643) of participants, with no one being excluded owing to additional criteria.

Ethics Approval

The survey was conducted in accordance with the Declaration of Helsinki, and the Ethics Committee of the Essen Medical Faculty (19-89-47-BO) agreed to conduct the study.

Measures

Overview

The survey contained items of sociodemographic, medical, and mental health data. In addition, we used a modified UTAUT questionnaire (based on previous adaptations) to assess the acceptance of e–mental health interventions and the resources of and barriers to eHealth use. The exact questionnaire is presented in Textbox 1. To assess the mental health of the participants, we used validated instruments such as the Eating Disorder Inventory-2–Bulimia (EDI-2–B), Eating Disorder Examination-Questionnaire 8 (EDE-Q8), and Patient Health Questionnaire-8 (PHQ-8).

Textbox 1. Scales and adapted items of Unified Theory of Acceptance and Use of Technology and references of original studies (italicized verbalizations have been adapted).

Behavioral intention (acceptance)

- "I would like to try a psychological online intervention." [26,40]

- "I would use a psychological online intervention if offered to me." [26,40]

- "I would recommend a psychological online intervention to my friends." [38]

Social influence

- "People close to me would approve the use of a psychological online intervention." [26,34,40]

- "My general practitioner would approve of a psychological online intervention." [26,40]

- "My friends would approve of a psychological online intervention." [21]

Performance expectancy

- "A psychological online intervention could improve my general well-being." [26,40]

- "A psychological online intervention could help me with stress." [26,40]

- "A psychological online intervention could help me improve my personal (psychological) health." [26,40]

Effort expectancy

- "The use of a psychological online intervention would not be an additional burden to me." (self-constructed)

- "A psychological online intervention would be easy to operate and comprehend." [26,33,40,41]

- "I could arrange using a

  psychological online intervention

  in my everyday life." [21]

Sociodemographic and Medical Data

Sociodemographic and medical data were assessed using items on age, sex, marital status, having children, occupational status, educational level, physical illness, mental disorder, and medication. In addition, there were items on data related to obesity and its management (weight, height, BMI, grade of obesity, comorbidities, and bariatric surgery).

Acceptance, eHealth Use, and UTAUT Predictors

To assess the acceptance of e–mental health interventions and its underlying factors, a modified version of the UTAUT model (Textbox 1) and several items for the measurement of internet use, internet anxiety, and attitudes toward and experiences with web-based interventions were used. The UTAUT questionnaire consists of 12 items and answers are given on a 5-point Likert scale (ranging from 1=totally disagree to 5=totally agree). Three items measure the underlying predictors of acceptance and acceptance itself, which are operationalized as intention to use (BI). Cronbach α values in this study were .88 for acceptance (BI), .84 for SI, .93 for PE, and .82 for EE, proving high internal consistency.

To assess the eHealth use of the participants, items such as the duration of use of media such their smartphone or tablet and previous experiences with eHealth interventions were asked. To record how confident the participants felt in using digital media, they were asked to rate their confidence on a scale of 1 (very unsafe) to 5 (very safe). The perceived stress caused by permanent availability via mobile phone or email was surveyed on a scale of 1 (strongly disagree) to 5 (strongly agree). The participants' internet anxiety was assessed using a set of 3 items (already used in previous studies), of which a mean value was calculated on a scale from 1 to 5, with 5 indicating very high internet anxiety. Cronbach α for this instrument in this study was .76, which indicates a sufficient internal consistency.

Assessment Using EDI-2–B

The EDI-2–B consists of 7 items assessing symptoms of bulimia (especially binge eating) on a 6-point Likert scale (1=never to 6=always) [42]. The sum score has a minimum of 7 points and a maximum of 42 points. Cronbach α in this study was .81, indicating high internal consistency.

Assessment Using EDE-Q8

The EDE-Q8 is a short version of the EDE-Q and comprises four subscales: restraint, eating concern, shape concern, and weight concern [43]. In this abbreviated version, 2 items refer to each scale (8 items in total), thus ensuring optimal internal consistency, one-dimensionality, and even coverage of the EDE-Q subscales. It consists of 5 items assessing eating disorder psychopathology in the past 28 days on a 7-point Likert scale (ranging from 0=not any day to 6=every day) and 3 items assessing the occurrence and frequency of core eating disorder behavior on a scale (from 0=never to 6=every time). Cronbach α in this study was .78, which indicates a sufficient internal consistency.

Assessment Using PHQ-8

The PHQ-8 measures depression symptoms via 8 items on a 4-point Likert scale (0=not at all to 3=nearly every day) [44]. A score ≥10 indicates major depression symptoms. Cronbach α in this study was .85, indicating high internal consistency.

Statistical Analyses

Data analysis was performed using SPSS Statistics 26 software (IBM). First, the sum scores and mean scores for scales PHQ-8, EDE-Q-8, and EDI-2–B were computed. Second, the internal consistencies for the different psychometric questionnaires were calculated and descriptive statistics were performed. Third, the acceptance was computed (mean value) and its distribution was assessed.

Acceptance (BI) of the UTAUT model (scale 1-5) was categorized by mean as low (1-2.34), moderate (2.35-3.67), and high (3.68-5) acceptance. Percentage and absolute count in categories were calculated. To highlight group differences in age, four age categories were formed before analysis: (1) 18-34 years, (2) 35-44 years, (3) 45-54 years, and (4) 55-69 years. The BMI of the participants was calculated by dividing their body weight by their height in meters squared. The means of acceptance (BI) were compared between groups regarding sociodemographic and medical data with 2-tailed t tests and analyses of variance (ANOVAs) to also include variables with multiple categories. The normal distribution of acceptance was examined using the Kolmogorov-Smirnov test, skewness, and kurtosis and graphically via a histogram including a normal distribution curve. All measures detected violations against normal distribution. However, we still used parametric tests for various reasons. According to the central limit theorem, the sampling distribution of the mean of a variable can be safely assumed to be normal if the variable and its mean are normally distributed in the population and the sample size is sufficiently large. We consider our sample size of 448 as sufficient because some studies suggest that such an effect already emerges at the sample size of 30 [45]. Moreover, other researchers found that acceptance distributions in general did not differ from normal distribution, which indicates that the variable acceptance might be normally distributed in the population [38]. In addition, 2-tailed t tests and ANOVAs are considered to be robust against violations, assuming normal distribution [46].

Using multiple hierarchical regression, the predictive model of acceptance was tested by using the enter method. The following predictors were included blockwise: (1) sociodemographic and medical data, (2) psychometric data, (3) eHealth-related data, and (4) UTAUT predictors (Figure 1). In addition, the full model was tested against the restricted UTAUT model with the UTAUT predictors (*PE, EE,* and *SI*) only. No multicollinearity could be detected because all the variance inflation factor values for testing multicollinearity were <5. The QQ plots of the residuals were visually inspected and showed no signs of violations against normality; therefore, normal distribution of the residuals can be assumed. Homoscedasticity was proven based on a scatter plot of the standardized residuals and adjusted predicted values.

For every ANOVA and 2-tailed t test, the level of significance was set at .05. In addition, post hoc tests were used to describe differences between the groups.

Results

Sociodemographic, Medical, and Psychometric Data

A total of 448 individuals participated in this study, of which 403 (89.9%) were women and 45 (10%) were men. The mean age was 44.69 years and age ranged from 18 to 69 years. Marital status for most participants was married (246/448, 54.9%) or living in a partnership (77/448, 17.2%). Of the 448 participants, 67 (14.9%) participants reported being single. Of the 448 participants, 98 (21.9%) participants had graduated from high school, 55 (12.3%) had graduated from college, and 210 (46.9%) had graduated from junior high school. In all, 66.3% (297/448) of the participants were employed at the time of participation and 33.7% (151/448) of the participants were unemployed.

A mental disorder was reported by 37.5% (168/448) of the participants. The diagnosis of depression was mentioned most frequently (58/168, 34.5%). There were 20.5% (92/448) of participants in outpatient psychotherapy owing to a mental disorder. In all, 22.8% (102/448) of the participants reported currently taking a psychiatric medication. The presence of a physical disease (other than obesity) was reported by 64.9% (291/448) of the participants. Frequently mentioned diseases (other than obesity) were diabetes, hypertension, and joint pain (182/448, 40.6%). Of the 448 participants with obesity, a total of 82 (18.3%) participants had obesity grade I, 88 (19.6%) participants had obesity grade II, and 278 (62.1%) participants had obesity grade III. Totally, 44.6% (200/448) of the participants had already undergone bariatric surgery. Of these 200 participants, 80 (40%) participants were part of the grade III obesity group. Of the 448 participants, 170 (37.9%) participants indicated that they planned to have bariatric surgery. Only 17.4% (78/448) participants neither had surgery nor planned to do so. Of the 200 participants who had already undergone surgery, 107 (53.5%) had opted for a gastric tube.

Of the 448 participants, 31 (6.9%) participants reported using the internet for private use for 0-1 hour per day, 94 (20.9%) participants used the internet for 1-2 hours per day, 119 (26.6%) participants reported a use time of 2-3 hours per day, 88 (19.6%) participants reported a use time of 3-4 hours per day, and approximately 116 (25.9%) people used the internet for >4 hours per day for private needs. The mean score for confidence in using digital media in this study was high (mean 4.08, SD 0.97). The stress experienced by the participants owing to permanent availability via mobile phone or email was moderate (mean 2.59, SD 1.24). Internet anxiety among participants in the study was low (mean 1.60, SD 0.76). Experience with e–mental health interventions was reported by 19.6% (88/448) of participants, whereas the remaining 80.4% (360/448) of the participants had no experience with such interventions.

The sum score of the EDI-2–B scale (mean 15.62, SD 0.87) was high compared with the norm values defined by Thiel and Paul [42], with a sample of 40.8% (183/448) of the participants (mean 11.59, SD 4.00) [43]. The sum score of EDE-Q8 (mean

3.89, SD 1.22) was also relatively high compared with healthy control patients from previous study by Hilbert et al [43] with a sample of 91.3% (409/448; mean 1.44, SD 1.22) and indicated eating disorder symptoms [44]. Using the recommended cutoff in the literature of ≥10, analyses of PHQ-8 (mean 10.10, SD 5.64) showed, that 49.6% (222/448) of the participants in this study might experience symptoms of major depression. In a very large random comparison sample (N=198,678) from a study in 2009, 8.6% of the participants experienced depressive symptoms (PHQ-8 sum score >10) [44].

Hypotheses 1 and 2: Acceptance as a Function of Sociodemographic and Medical Data

The general acceptance of e–mental health interventions was moderate (mean 3.18, SD 1.11). We divided the sample of 448 participants into groups by their degree of acceptance, with the results being as follows: 116 (25.9%) participants showed low acceptance, 199 (44.4%) showed moderate acceptance, and 133 (29.7%) showed high acceptance.

Table 1 contains the acceptance scores as a function of sociodemographic and medical data. Acceptance differed significantly between sexes (t_{446}=2.35; P=.02), with higher acceptance among women than among men. In addition, there were significant differences in acceptance by the different age groups ($F_{3,444}$=2.75; P=.04), with the highest acceptance in the middle-age group (35-44 years) and the lowest acceptance in the oldest group (55-69 years). However, the post hoc test of ANOVA showed that only age group II (35-44 years) and age group IV (55-69 years) differed significantly (P=.03). There was a significant difference in acceptance among the 3 obesity groups, with high acceptance level among people with obesity grade II ($F_{2,445}$=6.59; P=.002). The post hoc test of ANOVA revealed that the obesity grade I group was significantly different from grades II (P=.004) and III (P=.003). There was no significant difference in acceptance between grade II and grade III (P=.99). The occupational status (employed) was also significantly associated with higher acceptance ratings (t_{446}=2.40; P=.02) compared with those participants who were currently unemployed. In addition, participants with a mental disorder displayed significantly higher acceptance than those without a mental disorder (t_{446}=2.02; P=.02). There were no significant differences in acceptance regarding the variables such as bariatric surgery and outpatient psychotherapy.

Table 1. Differences in acceptance by sociodemographic and medical data (N=448).

Variable	Values, n (%)	Values, mean (SD)	t test (df)	F test (df)	P value
Sex			2.35 (446)	N/A[a]	.02
Women	403 (89.9)	3.23 (1.09)			
Men	45 (10)	2.78 (1.24)			
Age (years)			N/A	2.75 (3,444)	.04
18-34	86 (19.1)	3.15 (1.12)			
35-44	131 (29.2)	3.39 (1.12)			
45-54	140 (31.3)	3.16 (1.14)			
55-69	91 (20.3)	2.96 (1.03)			
Grade of obesity			N/A	6.59 (2,445)	.002
I	82 (18.3)	2.79 (1.06)			
II	88 (19.6)	3.34 (1.06)			
III	278 (62.1)	3.25 (1.12)			
Occupational status			2.40 (446)	N/A	.02
Employed	297 (66.3)	3.28 (1.11)			
Unemployed	151 (33.7)	3.01 (1.11)			
Mental disorder			2.02 (446)	N/A	.02
Yes	168 (37.5)	3.35 (1.05)			
No	280 (62.5)	3.09 (1.14)			
Bariatric surgery			N/A	1.04 (2,445)	.35
Yes, executed	200 (44.6)	3.11 (1.10)			
Already planned	170 (37.9)	3.28 (1.16)			
No	78 (17.4)	3.19 (1.07)			
Outpatient psychotherapy			0.795 (446)	N/A	.78
Yes	92 (20.5)	3.21 (1.16)			
No	356 (79.5)	3.11 (1.16)			

[a]N/A: not applicable.

Hypotheses 3 and 4: Predictors of Acceptance

The multiple hierarchical regression analysis revealed that the sociodemographic and medical predictors included in the first step explained 7.1% (R^2=0.071; $F_{5,442}$=6.74; P<.001) of the variance in acceptance. Thereby, sex (β=−0.13; P=.006), BMI (β=.13; P=.007), presence of a mental disorder (β=.16; P<.001), and occupational status (β=−0.18; P<.001) significantly predicted acceptance. The psychometric predictors included in the second step (R^2=0.127; $F_{8,439}$=7.98; P<.001) increased the explained variance significantly (ΔR^2=0.056; $F_{3,439}$=9.431; P<.001), whereby the depression sum score (PHQ-8; β=.12; P=.04) and the EDE-Q8 score (β=.13; P=.02) significantly predicted acceptance. In the third step, when entering the eHealth-related predictors, the explained variance in acceptance increased significantly to 15.4% (R^2=0.154; $F_{13,434}$=6.094; P<.001), but only the stress owing to constant availability on

the mobile phone or via email was a significant predictor in this step (β=.12; P=.02). The UTAUT predictors included in the last step (R^2=0.736; $F_{16,431}$=75.26; P<.001) changed the explained variance significantly by 58.2% (ΔR^2=0.582; $F_{3,431}$=317.28; P<.001), resulting in a total percentage of 73.6% for the explained variance. *PE* (β=.45; P<.001), *EE* (β=.22; P<.001), and *SI* (β=.27; P<.001) significantly predicted acceptance. In the overall model (step 4), in addition to the 3 UTAUT predictors, 3 additional variables from the extended model can be established as significant predictors of the acceptance of e–mental health interventions in people with obesity: *depression sum score* (PHQ-8; β=.07; P=.03), *stress owing to constant availability via mobile phone or* email (β=.06; P=.02), and *confidence in using digital media* (β=−0.06; P=.04). Table 2 presents the regression parameters of the hierarchical regression model of acceptance. Multimedia Appendix 1 illustrates the full model of the hierarchical regression.

XSL•FO

RenderX

Table 2. Hierarchical regression model of acceptance (N=448).

Predictor[a]	β[b]	B[c]	T	R^{2d}	ΔR^{2e}	P value
Step 1: sociodemographic and medical variables				0.071	0.071	
Sex	−0.130	−0.480	−2.788			.006
Age	−0.058	−0.006	−1.257			.21
Mental disorder	.156	.179	3.235			.001
BMI	.126	.017	2.760			.007
Occupational status	−0.182	−0.430	−3.760			<.001
Step 2: psychometric variables				0.127	0.056	
EDI-2–B[f]	.072	.013	1.384			.17
EDE-Q8[g]	.126	.115	2.293			.02
PHQ-8[h]	.124	.025	2.080			.04
Step 3: eHealth-related variables				0.154	0.027	
Stress owing to permanent availability	.112	.100	2.285			.02
Internet anxiety	−0.045	−0.067	−.884			.38
Information about eHealth interventions	.083	.088	1.664			.10
Experience with eHealth interventions	.092	.257	1.854			.06
Confidence in using digital media	.069	.079	1.368			.17
Step 4: UTAUT[i] predictors				0.736	0.582	
Performance expectancy	.454	.484	11.683			<.001
Effort expectancy	.219	.268	5.913			<.001
Social influence	.271	.361	7.956			<.001

[a]In steps 2, 3, and 4, only the newly included variables are presented.

[b]Standardized coefficient β.

[c]Unstandardized coefficient β.

[d]Determination coefficient.

[e]Changes in R^2.

[f]EDI-2–B: Eating Disorder Inventory-2–Bulimia.

[g]EDE-Q8: Eating Disorder Examination-Questionnaire 8.

[h]PHQ-8: Patient Health Questionnaire-8.

[i]UTAUT: Unified Theory of Acceptance and Use of Technology.

Hypothesis 5: UTAUT Versus Extended UTAUT Model

For hypothesis 5, we aimed to determine whether our full model (R^2=0.736) is better in explaining the variance in acceptance of e–mental health interventions than a restricted model (R^2=0.718), including only the UTAUT predictors of acceptance (*PE, EE,* and *SI*). Comparison of both models revealed a significant difference in the explained variance ($F_{13,431}$=2.366; *P*=.005), which means that the extended UTAUT model provides high variance explanation in the acceptance of e–mental health interventions owing to the additional included variables.

Discussion

Principal Findings

This study assessed the acceptance of e–mental health interventions among patients with obesity and explored the factors influencing acceptance. First, it is important to note that all hypotheses (1-5) were confirmed in this study. The overall acceptance of e–mental health interventions among people with obesity was moderate, with 29.7% (133/448) of participants indicating high acceptance, 44.4% (199/448) indicating moderate acceptance, and only 25.9% (116/448) indicating low acceptance of e–mental health interventions. Regarding the second hypothesis, the data yielded evidence for small but significant differences in acceptance depending on the obesity grade, with the highest acceptance in the grade II obesity group. Participants with obesity grade I differed significantly from those with obesity grades II and III in terms of their acceptance

of e–mental health interventions, with patients with obesity grade I exhibiting significantly low acceptance. Employed participants had a significantly higher acceptance of e–mental health interventions than unemployed participants, and participants with a mental disorder also showed significantly high acceptance; however, these differences were mostly slight. Regarding the third and fourth hypotheses, the following can be reported. In the extended regression model, the acceptance of e–mental health interventions was significantly predicted by depressive symptoms (PHQ-8). In addition, the stress caused by permanent availability via mobile phone or email was found to be a significant predictor of acceptance. Moreover, in the second step of the hierarchical regression, we used psychometric data that already explained a small part of the variance. This was different in previous research, which neglected the inclusion of psychometric data and only included eHealth-related data [21]. In addition to the 3 UTAUT predictors, confidence in using digital media was identified as a significant predictor in the regression model. Regarding the fifth hypothesis, the evidence is that the UTAUT predictors (restricted model with UTAUT predictors only) reached a high level in explained variance, whereas the extended UTAUT model (16 predictors) clarified slightly but with significantly more variance in patients with obesity (Table 2).

Comparison With Previous Work

Although previous studies have highlighted the mental health burden among patients with obesity, research regarding the acceptance and use of specific e–mental health interventions, especially with validated measures (eg, UTAUT), is still very scarce for the examined patient group [19]. The general acceptance of e–mental health interventions shown by participants was higher in this study than in previous studies [21,47,48], which also did not exclusively survey patients with obesity, who are a special patient group owing to their psychological and physical problems. We conducted a well-powered study comparable with previous research efforts, which often had a small sample, lacked valid measurement instruments, and captured very few variables that could be important for the acceptance of e–mental health interventions. We aimed to rectify these shortcomings, especially by using the validated measurement of acceptance of e–mental health interventions using the UTAUT model; adding sociodemographic, psychometric, and eHealth-related variables; and recruiting a substantial sample.

Studies that have previously examined the acceptance of e–mental health interventions in general or in specific patient groups have been able to identify the following variables to be significantly associated with acceptance: in addition to age [21,37,48], there is evidence of sex [37,48], anxiety [38], internet anxiety [38], experience with e–mental health interventions [21,37,47], education [21,37,48], experiencing a mental illness [37], and duration of type 2 diabetes [48]. Similar to previous research, the acceptance ratings in this study were significantly associated with sex. However, contrary to previous research by Hennemann et al [21] and Roelofsen et al [49], the acceptance ratings in this study were significantly higher for women than for men. We attribute this finding particularly to the topic of losing weight, dealing with their own bodies, and the

psychological factors in this regard, which more often plays a major role in the lives of women [50]. Age was also significantly associated with acceptance [21,49], whereby the middle-age group (35-44 years) differed significantly from the oldest group (55-69 years) in their acceptance, which could be because older people are less familiar with the internet and digital media. In addition, this study was successful in detecting the associations of the grade of obesity and experiencing a mental disorder with the acceptance of e–mental health interventions. According to previous research, it can be assumed that patients with a high BMI are more psychologically burdened; thus, it can be hypothesized that patient's acceptance of digital interventions also increases with higher weight (eg, owing to the immobility of the patient group), as these patients are more likely to be searching for psychological interventions and experiencing high levels of distress [51-53]. In this study, the level of distress and associated openness to psychological interventions among people with mental disorders also could possibly lead people who are currently experiencing a mental disorder to report high acceptance of e–mental health interventions. This could be in part because people with mental disorders are directly affected by the lack of psychosocial treatment possibilities and are more likely to be grateful to receive any low-threshold interventions to improve their symptoms. A practical implication that arises from this is the tailoring of specific eHealth interventions, especially to the psychological distress of these patients. Previous studies have used the UTAUT model to identify predictors of acceptance and use of internet-based interventions so that the following significant predictors could be identified: *PE* [21,48,54,55], *EE* [21,48,54,55], and *SI* [21,54,55].

The fourth core predictor of UTAUT has been named as *FCs* and is supposed to significantly predict the actual use. However, it does not predict the BI (or acceptance), which is why it was not included in our regression model [33]. The results of this study supported the viability of UTAUT in determining the acceptance of e–mental health interventions. The three UTAUT predictors, *PE*, *EE*, and *SI*, achieved a total of 71.8% at the variance explanation of the acceptance of e–mental health interventions, and this is comparable with those of the original UTAUT validation study (70%) [33]. This study found that *PE* was the key predictor of acceptance [21,33,54-56], which is consistent with previous research suggesting that *PE* is also a predictor of treatment outcome in psychotherapy [57]. The implications of the strong relationship between *PE* and acceptance of e–mental health interventions are that there must be transparent eHealth education in which misunderstandings or false expectations are openly addressed.

Beyond the 3 UTAUT predictors, previous research has identified the following factors as predictors of e–mental health acceptance: *perceived reliability* [55], *stress owing to permanent availability* [21], *perceived security* [48], *technology anxiety* [54], and *resistance to change* [54]. The overall model in this study (with 16 predictors) was significantly better than the restrictive UTAUT model, but ultimately explained only slightly more variance in acceptance. In this study, complementing previous research on this topic, we could also find the *sum score of the psychometric instrument PHQ-8* (indicator for symptoms of depression), the *confidence in using digital media* as

predictors of acceptance, and the *perceived stress through permanent availability*. In contrast to previous research, the regression coefficient of *perceived stress through permanent availability* in the overall model has a positive sign, which means that greater the stress perceived owing to constant accessibility, higher the acceptance was of e–mental health interventions. We explain this result as follows: people who report a high level of stress owing to permanent availability through their mobile phone use it very frequently, appreciate it, and use it for many different things in their everyday life. We assume that people who report a high level of stress from their mobile phone are particularly familiar with the functions and possibilities of their smartphone owing to the daily use. Therefore, they presumably exhibit a low inhibition threshold in the use of additional apps via smartphone and are more willing to use such technology for newly developed interventions or apps. People who reported low levels of stress from being available on their smartphone in the survey would also be unlikely to use their phone for e–mental health interventions because the smartphone does not have a significant role in their lives. This is a discovered difference from previous research, as, for example, Donkin and Glozier [58] identified technology fatigue as an important barrier to acceptance of e–mental health interventions and other findings, which also highlight that digital communication load is associated with psychological disorders such as burnout, anxiety, and depression [59]. *Confidence in using digital media* can be an influential variable concerning the acceptance of e–mental health interventions, as already shown by several studies related to internet use and internet literacy [60,61]. Again, the sign of the regression coefficient of *confidence in using digital media* is unexpected (negative), which, in interpretation first means that the more confident people are in using digital media, the less accepting they are of e–mental health interventions. However, when considering the entire regression model, it is noticeable that the sign of the regression coefficient for the predictor changes, which means that suppression effects might have occurred in the last step owing to the inclusion of the UTAUT factors. A practical implication resulting from this is to make as many patients as possible from different vulnerable patient groups familiar with the use of digital media and eHealth interventions so that they will be more widely accepted and used in the future.

Another new and important finding of this study is that psychometrics also contributes significantly to the variance explanation of acceptance of eHealth interventions among patients with obesity. Therefore, as a theoretical implication, this result should be included in the analysis of future research. The depression score (PHQ-8) is a significant predictor of the acceptance of e–mental health interventions in the overall regression model. As discussed above regarding mental disorders, we can assume that people with high depression scores and more prominent psychological symptoms generally experience more distress. This can lead to the finding that the acceptance of e–mental health interventions seems to be high among patients who are currently experiencing psychological distress.

The evaluation of the psychometric data (EDI-2–B, EDE-Q8, and PHQ-8) shows that the current sample of patients with obesity is noticeably psychologically burdened. The two kinds of psychological symptoms are eating disorders, such as binge eating, and depressive symptoms. A large review of obesity and psychiatric disorders from 2017 shows a very strong association between obesity and depression, especially in longitudinal studies, where the correlation was stronger for women. In addition, multiple studies have also shown an association between eating disorder symptoms and obesity [8]. This leads to the practical implication that eHealth interventions should be particularly targeted at restoring mental health, especially in patients with obesity.

In general, this study confirms recent findings (owing to high variance explanation of the 3 UTAUT factors in the acceptance of e–mental health interventions) and supplements them with further predictors (*depression, stress caused by permanent availability,* and *confidence in using digital media*).

However, it can be assumed that further factors influence the results of acceptance, especially the use of e–mental health interventions, and these must be taken into account when looking retrospectively at the results of previous research. In the evaluation of previous findings on acceptance and use of e–mental health interventions, we should consider whether the researchers used the UTAUT model or another measurement instrument to describe acceptance, the type and duration of the patient's illness, the type of eHealth service, and whether patients or health care workers were surveyed.

Limitations

The following limitations should be considered when interpreting the results of this study. As our study was exclusively web-based, it was mandatory that participants had internet access. As the spread of internet access varies, particularly between age groups, our sample of patients with obesity tends to be younger than the average in the general population (only 5 people older than 65 years). Moreover, the sex distribution is not representative of the overall population. As we recruited participants widely through social media groups (obesity surgery–related groups, which are almost exclusively composed of female members), a large number of women participated (403/448, 89.9% women and 45/448, 10% men), which restricts the generalizability of the results. In addition, we had varying numbers of patients in each obesity grade group: obesity grade I (82/448, 18.3%), obesity grade II (88/448, 19.6%), and obesity grade III (278/448, 62.1%). Thus, a significantly high number of persons are in obesity grade III, which also does not correspond to the distribution in society [62]. The BMI bias in this study (with 278/448, 62.1% of the participants with a BMI >40 kg/m^2) is particularly important, as eHealth interventions often have preventive functions and could be valuable for patients in obesity groups I or II. Owing to these sampling biases, the generalizability of our results may once again be reduced. In addition to the unbalanced distribution of participants from the different obesity groups, we recruited a very high number of participants who had already undergone bariatric surgery (200/448, 44.6%) or were in the process of planning to undergo the surgery (170/448, 37.9%), which could

also limit the generalizability of our results. Similarly, a considerable number of participants did not have a diagnosed mental health disorder at the time of the survey (which makes comparability difficult), whereby acceptance of e–mental health interventions was high among individuals with a mental health disorder. Nevertheless, the general psychological distress of the participants in this study was high, as measured by the valid psychometric instruments.

Owing to the recruitment of participants via the internet (in particular, via social media), we can assume that they might already have been more willing and interested in internet-related topics than random participants from different social backgrounds; therefore, we cannot rule out a selection bias. It is important to note that all the data collected were self-reported. Thus, accuracy of the results may be limited by the fact that participants may respond in a very socially elicited manner. Self-reporting can lead to a phenomenon known as common method bias [63]. To counteract this limitation, the instruments used in the study had sufficient reliability, the survey had a defensible length and was anonymous and web-based, and the patients were well educated, as these points are known to mitigate common method bias.

It is important to mention that this study, similar to most of the previous studies, determined the acceptance of e–mental health interventions among patients with obesity only by using the BI. A direct inference from the intention to use an e–mental health intervention to the actual use is not possible owing to the intention–behavior gap. However, further research should take the limitations of this study into account and include the actual use (behavior) of e–mental health interventions in patients with obesity and not focus only on acceptance.

As a theoretical implication for future research that also focuses on capturing the acceptance of e–mental health interventions, it would be particularly important to observe the distribution of sex, age, and the different obesity groups, to keep it as representative as possible. In addition, it would be beneficial to conduct a longitudinal study in which barriers and predictors for the actual use behavior of e–mental health services would be identified because no causality can be determined by cross-sectional studies. In addition, it would certainly be conceivable to survey other special patient groups, who, similar to patients with obesity (owing to stigma, physical illness, comorbidities, and immobility), have specific barriers that make the implementation of e–mental health interventions particularly important. This would facilitate the identification of the specific needs and demands of these patient groups and, as a practical implication, the development and implementation of e–health interventions that specifically target the improvement of mental and physical health.

Conclusions

Although the measured acceptance in patients with obesity could be determined as moderate, this study highlights that the acceptance of e–mental health interventions differs significantly depending on the following variables: age, grade of obesity, occupational status, sex, and mental health status.

The UTAUT model with its three core predictors (*PE*, *EE*, and *SI*) has proven to be a valuable instrument to predict the acceptance of e–mental health interventions in patients with obesity. The variance explained by acceptance in the restrictive UTAUT model (the 3 core predictors) was high, but the extended UTAUT model is slightly but significantly better in comparison and highlighted three additional significant predictors (depression, stress owing to constant availability via mobile phone or email, and confidence in using digital media). Owing to the close association between acceptance and use, acceptance-facilitating interventions should be fostered to enhance the establishment of effective e–mental health interventions for patients with obesity. Low-threshold, location-flexible, and efficient e–mental health interventions are more important than ever before, especially with regard to the ongoing pandemic and in light of the high psychological vulnerability of patients with obesity. Especially because many patients with obesity decide to undergo bariatric surgery, such interventions could be very relevant in the preoperative phase (for psychological support) and in the postoperative follow-up. In addition, further research should be conducted to determine the detailed expectations, needs, and demands of patients with obesity regarding such tailor-made interventions to further increase motivation and acceptance.

Authors' Contributions

AB, MT, EMS, and MN initiated and conceptualized the study. LCS was coresponsible for the recruitment of the participants. VR, MD, JS, and AS performed the statistical analyses and interpretation of the data, and VR wrote the first draft of the manuscript. VR, MD, and AS performed the data acquisition and statistical analyses. AB, MT, and EMS contributed to the design of the study. All authors contributed to the further writing of the manuscript and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full hierarchical regression model of acceptance.
[DOCX File , 20 KB - formative_v6i3e31229_app1.docx]

References

1.   Obesity and overweight. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/
      obesity-and-overweight [accessed 2022-03-07]
2.   Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of
      overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease
      Study 2013. Lancet 2014 Aug 30;384(9945):766-781 [FREE Full text] [doi: 10.1016/S0140-6736(14)60460-8] [Medline:
      24880830]
3.   Puigdomenech Puig E, Robles N, Saigí-Rubió F, Zamora A, Moharra M, Paluzie G, et al. Assessment of the efficacy, safety,
      and effectiveness of weight control and obesity management mobile health interventions: systematic review. JMIR Mhealth
      Uhealth 2019 Oct 25;7(10):e12612 [FREE Full text] [doi: 10.2196/12612] [Medline: 31654566]
4.   Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mHealth interventions
      for obesity and diabetes treatment and self-management. Adv Nutr 2017 May;8(3):449-462 [FREE Full text] [doi:
      10.3945/an.116.014100] [Medline: 28507010]
5.   Zizza C, Herring A, Stevens J, Popkin B. Obesity affects nursing-care facility admission among whites but not blacks. Obes
      Res 2002 Aug;10(8):816-823 [FREE Full text] [doi: 10.1038/oby.2002.110] [Medline: 12181391]
6.   Löbner M, Stein J, Rost T, Förster F, Dorow M, Keller J, et al. [innovative e-health approaches for comorbid depression
      in patients with obesity: patient and expert perspectives on user acceptance]. Psychiatr Prax 2017 Jul;44(5):286-295. [doi:
      10.1055/s-0043-107471] [Medline: 28403502]
7.   Avila C, Holloway AC, Hahn MK, Morrison KM, Restivo M, Anglin R, et al. An overview of links between obesity and
      mental health. Curr Obes Rep 2015 Sep;4(3):303-310. [doi: 10.1007/s13679-015-0164-9] [Medline: 26627487]
8.   Rajan T, Menon V. Psychiatric disorders and obesity: a review of association studies. J Postgrad Med 2017;63(3):182-190
      [FREE Full text] [doi: 10.4103/jpgm.JPGM_712_16] [Medline: 28695871]
9.   Taylor VH, Forhan M, Vigod SN, McIntyre RS, Morrison KM. The impact of obesity on quality of life. Best Pract Res
      Clin Endocrinol Metab 2013 Apr;27(2):139-146. [doi: 10.1016/j.beem.2013.04.004] [Medline: 23731876]
10.   Kinzl JF, Maier C, Bösch A. [Morbidly obese patients: psychopathology and eating disorders. Results of a preoperative
      evaluation]. Neuropsychiatr 2012;26(4):159-165. [doi: 10.1007/s40211-012-0036-4] [Medline: 23179359]
11.   Mannan M, Mamun A, Doi S, Clavarino A. Prospective associations between depression and obesity for adolescent males
      and females- a systematic review and meta-analysis of longitudinal studies. PLoS One 2016;11(6):e0157240 [FREE Full
      text] [doi: 10.1371/journal.pone.0157240] [Medline: 27285386]
12.   Schwenke M, Löbner M, Riedel-Heller S, Luppa M. [Obesity and depression in primary care - results from the INTERACT
      study]. Psychiatr Prax 2020 Oct;47(7):388-391. [doi: 10.1055/a-1144-7035] [Medline: 32268415]
13.   Puhl RM, Heuer CA. Obesity stigma: important considerations for public health. Am J Public Health 2010
      Jun;100(6):1019-1028. [doi: 10.2105/AJPH.2009.159491] [Medline: 20075322]
14.   Abdelaal M, le Roux CW, Docherty NG. Morbidity and mortality associated with obesity. Ann Transl Med 2017 Apr;5(7):161
      [FREE Full text] [doi: 10.21037/atm.2017.03.107] [Medline: 28480197]
15.   Batsis JA, Pletcher SN, Stahl JE. Telemedicine and primary care obesity management in rural areas - innovative approach
      for older adults? BMC Geriatr 2017 Jan 05;17(1):6 [FREE Full text] [doi: 10.1186/s12877-016-0396-x] [Medline: 28056832]
16.   Bäuerle A, Teufel M, Musche V, Weismüller B, Kohler H, Hetkamp M, et al. Increased generalized anxiety, depression
      and distress during the COVID-19 pandemic: a cross-sectional study in Germany. J Public Health (Oxf) 2020 Nov
      23;42(4):672-678 [FREE Full text] [doi: 10.1093/pubmed/fdaa106] [Medline: 32657323]
17.   Bäuerle A, Graf J, Jansen C, Musche V, Schweda A, Hetkamp M, et al. E-mental health mindfulness-based and skills-based
      'CoPE It' intervention to reduce psychological distress in times of COVID-19: study protocol for a bicentre longitudinal
      study. BMJ Open 2020 Aug 13;10(8):e039646 [FREE Full text] [doi: 10.1136/bmjopen-2020-039646] [Medline: 32792455]
18.   Eccleston C, Blyth FM, Dear BF, Fisher EA, Keefe FJ, Lynch ME, et al. Managing patients with chronic pain during the
      COVID-19 outbreak: considerations for the rapid introduction of remotely supported (eHealth) pain management services.
      Pain 2020 May;161(5):889-893 [FREE Full text] [doi: 10.1097/j.pain.0000000000001885] [Medline: 32251203]
19.   Hedman E, Andersson E, Ljótsson B, Andersson G, Rück C, Lindefors N. Cost-effectiveness of internet-based cognitive
      behavior therapy vs. cognitive behavioral group therapy for social anxiety disorder: results from a randomized controlled
      trial. Behav Res Ther 2011 Nov;49(11):729-736. [doi: 10.1016/j.brat.2011.07.009] [Medline: 21851929]
20.   Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression
      disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78
      [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
21.   Hennemann S, Beutel ME, Zwerenz R. Drivers and barriers to acceptance of web-based aftercare of patients in inpatient
      routine care: a cross-sectional survey. J Med Internet Res 2016 Dec 23;18(12):e337 [FREE Full text] [doi: 10.2196/jmir.6003]
      [Medline: 28011445]
22.   Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior
      therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014
      Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]

23. Messiah SE, Sacher PM, Yudkin J, Ofori A, Qureshi FG, Schneider B, et al. Application and effectiveness of eHealth strategies for metabolic and bariatric surgery patients: a systematic review. Digit Health 2020;6:2055207619898987 [FREE Full text] [doi: 10.1177/2055207619898987] [Medline: 32030193]

24. Wright C, Mutsekwa RN, Hamilton K, Campbell KL, Kelly J. Are eHealth interventions for adults who are scheduled for or have undergone bariatric surgery as effective as usual care? A systematic review. Surg Obes Relat Dis 2021 Dec;17(12):2065-2080. [doi: 10.1016/j.soard.2021.07.020] [Medline: 34474983]

25. Eichenberg C, Ott R. Klinisch-psychologische intervention im internet. Psychotherapeut 2011 Jun 10;57(1):58-69 [FREE Full text] [doi: 10.1007/s00278-011-0832-5]

26. Ebert D, Berking M, Cuijpers P, Lehr D, Pörtner M, Baumeister H. Increasing the acceptance of internet-based mental health interventions in primary care patients with depressive symptoms. A randomized controlled trial. J Affect Disord 2015 May 01;176:9-17. [doi: 10.1016/j.jad.2015.01.056] [Medline: 25682378]

27. Berry D, Blonquist T, Patel R, Halpenny B, McReynolds J. Exposure to a patient-centered, web-based intervention for managing cancer symptom and quality of life issues: impact on symptom distress. J Med Internet Res 2015 Jun 03;17(6):e136 [FREE Full text] [doi: 10.2196/jmir.4190] [Medline: 26041682]

28. Boonstra A, Broekhuis M. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. BMC Health Serv Res 2010 Aug 06;10:231 [FREE Full text] [doi: 10.1186/1472-6963-10-231] [Medline: 20691097]

29. Sinclair C, Holloway K, Riley G, Auret K. Online mental health resources in rural Australia: clinician perceptions of acceptability. J Med Internet Res 2013 Sep 05;15(9):e193 [FREE Full text] [doi: 10.2196/jmir.2772] [Medline: 24007949]

30. Hennemann S, Beutel ME, Zwerenz R. Ready for eHealth? Health professionals' acceptance and adoption of eHealth interventions in inpatient routine care. J Health Commun 2017 Mar;22(3):274-284. [doi: 10.1080/10810730.2017.1284286] [Medline: 28248626]

31. Eichenberg C, Wolters C, Brähler E. The internet as a mental health advisor in Germany--results of a national survey. PLoS One 2013;8(11):e79206 [FREE Full text] [doi: 10.1371/journal.pone.0079206] [Medline: 24278121]

32. Currie M, Philip L, Roberts A. Attitudes towards the use and acceptance of eHealth technologies: a case study of older adults living with chronic pain and implications for rural healthcare. BMC health services research 2015 Apr 16;15:162 [FREE Full text] [doi: 10.1186/s12913-015-0825-0] [Medline: 25888988]

33. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319-340. [doi: 10.2307/249008]

34. Venkatesh V, Morris M, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425-478. [doi: 10.2307/30036540]

35. Senft N, Abrams J, Katz A, Barnes C, Charbonneau DH, Beebe-Dimmer JL, et al. eHealth activity among African American and white cancer survivors: a new application of theory. Health Commun 2020 Mar;35(3):350-355 [FREE Full text] [doi: 10.1080/10410236.2018.1563031] [Medline: 32013612]

36. de Veer AJ, Peeters J, Brabers A, Schellevis FG, Rademakers JJ, Francke AL. Determinants of the intention to use e-Health by community dwelling older people. BMC Health Serv Res 2015 Mar 15;15:103 [FREE Full text] [doi: 10.1186/s12913-015-0765-8] [Medline: 25889884]

37. Crisp DA, Griffiths KM. Participating in online mental health interventions: who is most likely to sign up and why? Depress Res Treat 2014;2014:790457 [FREE Full text] [doi: 10.1155/2014/790457] [Medline: 24804089]

38. Lin J, Faust B, Ebert DD, Krämer L, Baumeister H. A web-based acceptance-facilitating intervention for identifying patients' acceptance, uptake, and adherence of internet- and mobile-based pain interventions: randomized controlled trial. J Med Internet Res 2018 Aug 21;20(8):e244 [FREE Full text] [doi: 10.2196/jmir.9925] [Medline: 30131313]

39. World Health Organization. Obesity Preventing and Managing the Global Epidemic. Geneva, Switzerland: World Health Organization; 2000.

40. Baumeister H, Nowoczin L, Lin J, Seifferth H, Seufert J, Laubner K, et al. Impact of an acceptance facilitating intervention on diabetes patients' acceptance of internet-based interventions for depression: a randomized controlled trial. Diabetes Res Clin Pract 2014 Jul;105(1):30-39. [doi: 10.1016/j.diabres.2014.04.031] [Medline: 24862240]

41. Liu L, Miguel Cruz A, Rios Rincon A, Buttar V, Ranson Q, Goertzen D. What factors determine therapists' acceptance of new technologies for rehabilitation – a study using the Unified Theory of Acceptance and Use of Technology (UTAUT). Disabil Rehabil 2015;37(5):447-455. [doi: 10.3109/09638288.2014.923529] [Medline: 24901351]

42. Thiel A, Paul T. Entwicklung einer deutschsprachigen Version des Eating-Disorder-Inventory (EDI) [Development of a German-language version of the Eating Disorder Inventory (EDI)]. Zeitschrift für Differentielle und Diagnostische Psychologie 1988;9(4):267-278.

43. Hilbert A, Tuschen-Caffier B, Karwautz A, Niederhofer H, Munsch S. Eating disorder examination-questionnaire. Diagnostica 2007 Jul;53(3):144-154. [doi: 10.1026/0012-1924.53.3.144]

44. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]

45. The assumption(s) of normality. Mordkoff JT. URL: http://www2.psychology.uiowa.edu/faculty/mordkoff/GradStats/part%201/I.07%20normal.pdf [accessed 2022-03-07]

46. Wilcox R. Chapter 1 - Introduction. In: Introduction to Robust Estimation and Hypothesis Testing (Third Edition). Cambridge, Massachusetts: Academic Press; 2012.

47. Gun SY, Titov N, Andrews G. Acceptability of internet treatment of anxiety and depression. Australasian Psychiatry 2011 Jun;19(3):259-264. [doi: 10.3109/10398562.2011.562295] [Medline: 21682626]

48. Cimperman M, Makovec Brenčič M, Trkman P. Analyzing older users' home telehealth services acceptance behavior-applying an Extended UTAUT model. Int J Med Inform 2016 Jun;90:22-31. [doi: 10.1016/j.ijmedinf.2016.03.002] [Medline: 27103194]

49. Roelofsen Y, Hendriks SH, Sieverink F, Landman GW, Groenier KH, Bilo HJ, et al. Differences between patients with type 2 diabetes mellitus interested and uninterested in the use of a patient platform (e-vitadm-2/zodiac-41). J Diabetes Sci Technol 2014 Mar;8(2):230-237 [FREE Full text] [doi: 10.1177/1932296814524496] [Medline: 24876572]

50. Silberstein L, Striegel-Moore R, Timko C, Rodin J. Behavioral and psychological implications of body dissatisfaction: do men and women differ? Sex Roles 1988 Aug;19(3-4):219-232 [FREE Full text] [doi: 10.1007/bf00290156]

51. Herhaus B, Kersting A, Brähler E, Petrowski K. Depression, anxiety and health status across different BMI classes: a representative study in Germany. J Affect Disord 2020 Nov 01;276:45-52. [doi: 10.1016/j.jad.2020.07.020] [Medline: 32697715]

52. Luppino FS, de Wit L, Bouvy PF, Stijnen T, Cuijpers P, Penninx BW, et al. Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. Arch Gen Psychiatry 2010 Mar;67(3):220-229. [doi: 10.1001/archgenpsychiatry.2010.2] [Medline: 20194822]

53. Sockalingam S, Leung SE, Cassin SE. The impact of coronavirus disease 2019 on bariatric surgery: redefining psychosocial care. Obesity (Silver Spring) 2020 Jun;28(6):1010-1012 [FREE Full text] [doi: 10.1002/oby.22836] [Medline: 32294297]

54. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: an extension of the UTAUT model. Int J Med Inform 2017 May;101:75-84. [doi: 10.1016/j.ijmedinf.2017.02.002] [Medline: 28347450]

55. Alam M, Wang H, Zapan B. Using the UTAUT model to determine factors affecting acceptance and use of mobile health (mHealth) services in Bangladesh. J Stud Soc Sci 2018;17(2):137-172 [FREE Full text]

56. Taiwo A, Downe AG. The theory of user acceptance and use of technology (UTAUT): a meta-analytic review of empirical findings. J Theoretical Applied Inf Technol 2013;49(1):48-58 [FREE Full text]

57. Constantino MJ, Arnkoff DB, Glass CR, Ametrano RM, Smith JZ. Expectations. J Clin Psychol 2011 Feb;67(2):184-192. [doi: 10.1002/jclp.20754] [Medline: 21128304]

58. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012 Jun 22;14(3):e91 [FREE Full text] [doi: 10.2196/jmir.2100] [Medline: 22743581]

59. Reinecke L, Aufenanger S, Beutel ME, Dreier M, Quiring O, Stark B, et al. Digital stress over the life span: the effects of communication load and internet multitasking on perceived stress and psychological health impairments in a German probability sample. Media Psychol 2016 Mar 21;20(1):90-115. [doi: 10.1080/15213269.2015.1121832]

60. Frees B, Koch W. Internetnutzung: Frequenz und Vielfalt nehmen in allen Altersgruppen zu. Media Perspektiven 9/2015. 2015. URL: https://www.ard-zdf-onlinestudie.de/files/2015/0915_Frees_Koch.pdf [accessed 2022-03-07]

61. van Deursen AJ, van Dijk JA, Peters O. Rethinking internet skills: the contribution of gender, age, education, internet experience, and hours online to medium- and content-related internet skills. Poetics 2011 Apr;39(2):125-144. [doi: 10.1016/j.poetic.2011.02.001]

62. Schienkiewitz A, Mensink G, Kuhnert R, Lange C. Übergewicht und Adipositas bei Erwachsenen in Deutschland. GBE-BUND. URL: https://www.gbe-bund.de/gbe/abrechnung.prc_abr_test_logon?p_uid=gast&p_aid=0&p_knoten=FID&p_sprache=D&p_suchstring=25541 [accessed 2022-03-07]

63. Conway J, Lance C. What reviewers should expect from authors regarding common method bias in organizational research. J Bus Psychol 2010 May 22;25(3):325-334 [FREE Full text] [doi: 10.1007/s10869-010-9181-6]

Abbreviations

ANOVA: analysis of variance
BI: behavioral intention
EDE-Q8: Eating Disorder Examination-Questionnaire 8
EDI-2–B: Eating Disorder Inventory-2–Bulimia
EE: effort expectancy
FC: facilitating condition
PE: performance expectancy
PHQ-8: Patient Health Questionnaire-8
SI: social influence
UTAUT: Unified Theory of Acceptance and Use of Technology

XSL•FO
RenderX

<u>Original Paper</u>

Testing Digital Methods of Patient-Reported Outcomes Data Collection: Prospective Cluster Randomized Trial to Test SMS Text Messaging and Mobile Surveys

Anish K Agarwal[1,2,3], MD, MPH, MS; Zarina S Ali[2,4], MD; Frances Shofer[1], PhD; Ruiying Xiong[5], MS; Jessica Hemmons[1,5], MS; Evan Spencer[1,5], BS; Dina Abdel-Rahman[1,5], BS; Brian Sennett[6], MD; Mucio K Delgado[1,2,5], MD, MS

[1]Department of Emergency Medicine, University of Pennsylvania, Philadelphia, PA, United States

[2]Leonard Davis Institute, University of Pennsylvania, Philadelphia, PA, United States

[3]Penn Medicine Center for Health Care Innovation, University of Pennsylvania, Philadelphia, PA, United States

[4]Department of Neurosurgery, University of Pennsylvania, Philadelphia, PA, United States

[5]Behavioral Science and Analytics for Injury Reduction, University of Pennsylvania, Philadelphia, PA, United States

[6]Department of Orthopedic Surgery, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Anish K Agarwal, MD, MPH, MS
Department of Emergency Medicine
University of Pennsylvania
3400 Spruce Street
Philadelphia, PA, 19104
United States
Phone: 1 6103042318
Email: anish.agarwal@pennmedicine.upenn.edu

Abstract

Background: Health care delivery continues to evolve, with an effort being made to create patient-centered care models using patient-reported outcomes (PROs) data. Collecting PROs has remained challenging and an expanding landscape of digital health offers a variety of methods to engage patients.

Objective: The aim of this study is to prospectively investigate two common methods of remote PRO data collection. The study sought to compare response and engagement rates for bidirectional SMS text messaging and mobile surveys following orthopedic surgery.

Methods: The study was a prospective, block randomized trial of adults undergoing elective orthopedic procedures over 6 weeks. The primary objective was to determine if the method of digital patient engagement would impact response and completion rates. The primary outcome was response rate and total completion of PRO questionnaires.

Results: A total of 127 participants were block randomized into receiving a mobile survey (n=63) delivered as a hyperlink or responding to the same questions through an automated bidirectional SMS text messaging system (n=64). Gender, age, number of comorbidities, and opioid prescriptions were similar across messaging arms. Patients receiving the mobile survey were more likely to have had a knee-related surgery (n=50, 83.3% vs n=40, 62.5%; *P*=.02) but less likely to have had an invasive procedure (n=26, 41.3% vs n=39, 60.9%; *P*=.03). Overall engagement over the immediate postoperative period was similar. Prolonged engagement for patients taking opioids past postoperative day 4 was higher in the mobile survey arm at day 7 (18/19, 94.7% vs 9/16, 56.3%). Patients with more invasive procedures showed a trend toward being responsive at day 4 as compared to not responding (n=41, 59.4% vs n=24, 41.4%; *P*=.05).

Conclusions: As mobile patient engagement becomes more common in health care, testing the various options to engage patients to gather data is crucial to inform future care and research. We found that bidirectional SMS text messaging and mobile surveys were comparable in response and engagement rates; however, mobile surveys may trend toward higher response rates over longer periods of time.

Trial Registration: ClinicalTrials.gov NCT03532256; https://clinicaltrials.gov/ct2/show/NCT03532256

XSL•FO
RenderX

Introduction

Health care delivery continues to evolve, with an effort being made to create patient-centered care models. Learning health systems engage patients and elicit patient-reported outcomes (PROs) data to continuously improve and drive clinical practice [1,2]. Incorporating PROs into clinical decision-making has led to improvements in patients' quality of life, improved communication, and reductions in unscheduled care [3-5]. Collecting PROs has remained challenging and their integration into clinical decisions remains understudied [6].

The use of mobile technology (cell phones, smartphones, tablets) continues to increase in society and health [7,8]. Engaging patients to capture PROs may provide clinicians with additional understanding to support care and motivate behavior remotely [9,10]. As digital strategies continue to expand, there has been limited research on the optimal ways to reach patients and capture PROs. Research using mobile health has grown in the past decade [11-13]. Clinicians have attempted to use mobile technology to track medication adherence [14,15], encourage healthy behaviors [16], improve home monitoring [17], and institute automated "hovering" to track chronic diseases [18]. However, less is known about the various mobile methods of engaging with patients and collecting and monitoring patient-reported data.

Within the context of the opioid epidemic and in an effort to promote opioid prescribing stewardship, surgical studies have collected PROs focused on pain intensity and opioid prescribing and use through surveys to reduce excessive prescribing [19-22]. These studies have revealed a mismatch between prescribing and patient-reported use, but are limited by their retrospective design, recall bias, and limited response rates. PROs have been used in orthopedic surgery to help guide preoperative decision-making and improve patient satisfaction [9,23]. To overcome the limitations of prior research, the existing gap regarding prospective PRO data on pain, function, and opioid consumption must be addressed. The rapid evolution of mobile technology may provide an opportunity for clinicians to assess trends in patients' self-reported pain and use of prescription analgesics, test methods to support safe and effective pain management, and translate data into clinical protocols [24].

The purpose of this study was to prospectively investigate two mobile patient engagement strategies to collect PRO data. Surgeons seek PRO data on acute pain management to inform safe opioid prescribing and to effectively manage acute pain. We compared response and completion rates of postoperative PROs among patients using bidirectional SMS text messaging versus mobile hyperlink surveys. To our knowledge, this is the first study to compare two distinct mobile engagement approaches in this context. The hypothesis of this study was that conversational SMS text messaging would result in higher response and completion rates as compared to mobile survey hyperlinks.

Methods

Overview

This was a prospective, block randomized trial of adult patients (aged 18 years or older) undergoing elective orthopedic procedures (ClinicalTrials.gov NCT03532256). The study took place over 6 weeks between July 1, 2019, and August 12, 2019. The primary objective was to determine if the method of patient engagement would impact response and completion rates. The primary outcome was response rate and total completion of PRO questionnaires.

Eligible patients included adult patients undergoing an elective outpatient orthopedic surgery (including knee, hip, shoulder, and elbow) and prescribed an acute opioid for postoperative pain. Acute opioid pain medications included oxycodone, oxycodone/acetaminophen, hydromorphone, or hydrocodone. Exclusion criteria included no access to an SMS-capable device or no opioid prescription. These inclusion and exclusion criteria were consistent with prior published research protocols using an established automated postoperative messaging program [24,25]. All patients were recruited from the University of Pennsylvania Health System Department of Orthopedics, which performs approximately 13,000 surgeries per year. In addition to general procedures (~3500 annual surgeries), the department consists of the following divisions: (1) Sports Medicine (~1700 annual surgeries), (2) Hand (~1900), (3) Joints (~4500), and (4) Trauma (~1500).

The research team had previously worked with the institution's legal, privacy, and patient safety departments to obtain remote, SMS text message–based consent for data collection [25]. This was an intentional design of a larger institutional program aimed at improving acute opioid prescribing. The approach allows for remote consent in an attempt to reduce clinical providers' workload and improve the scale of engaging eligible patients.

Following an elective outpatient surgery within the Department of Orthopedics, patients who underwent surgery received an initial SMS text message on the second day after the procedure. This message informed the patient of safe SMS text messaging data practices, provided links to further information regarding the follow-up research study, and offered the ability to opt in or opt out of further messaging. Patient were asked to consent via a simple SMS text message response of "yes." Consenting patients were then sent SMS text messages on postoperative days number 4, 7, 14, 21, and 28. Patients were block randomized in groups of 2 or 4 to receive either (1) a hyperlink to a web-based mobile survey or (2) automated bidirectional SMS text messaging in a conversational format. Mobile survey questions and conversational questions were identical and included PROs related to self-reported pain intensity, ability to

manage pain, use of prescribed medications, and ability to control pain. Block randomization was used to achieve balance in the two groups across demographics, procedure date, and dates of mobile engagement.

Individuals who either did not respond, opted out, or self-reported no further planned use of acute opioid medications were not subsequently messaged. For example, if a participant completed the mobile survey or replied to the bidirectional SMS text messaging on postoperative day 4 and indicated they were no longer planning on using their opioid prescription, no further messaging was sent on day 7, 14, 21, or 28. Participants who did not respond to any messaging would receive a reminder message within 30 minutes; if participants did not respond, then no further messaging was sent. The follow-up intervals were determined with input from a key clinician and surgeon and were in line with prior studies that evaluated PROs at 1 or 2 weeks postoperation.

Descriptive summary statistics were used to characterize the study population, using mean and standard deviation for age, and frequencies and percentages for categorical variables including gender, type of surgery, opioid tablet quantity, comorbidities, and invasiveness of the procedure (open surgical approach vs laparoscopic). To determine differences in the primary outcome of response rate at day 4 between the messaging arms and categorical demographic variables, Fisher exact test was used. To determine the difference in response at day 4 by age, Student t test was used. To determine if there were differences in secondary outcomes such as opioid use over time, Kaplan-Meier product-limit method with a log-rank test was used. A 2-sided α of .025 was considered statistically significant. All opioid prescription types were converted to equivalent doses of 5 mg oxycodone tablets [26]. All analyses were performed using SAS statistical software (version 9.4; SAS Institute).

Ethics Approval

The Institutional Review Board of the University of Pennsylvania approved this study (number 827461).

Results

Over 6 weeks, 127 participants were block randomized into either receiving a mobile survey (n=63) delivered as a hyperlink or responding to the same questions through an automated system—a chatbot using bidirectional SMS text messaging (n=64). Overall, 57.4% (n=73) of participants were male, the mean age was 38.5 (SD 13.9) years, and 70.8% (n=90) had had a knee procedure. Gender, age, number of comorbidities, and number of opioid tablets prescribed were similar across arms (Table 1). Patients receiving the mobile survey were more likely to have had a knee procedure (n=50, 83.3% vs n=40, 62.5%; P=.02) but less likely to have had an invasive or open procedure (n=26, 41.3% vs n=39, 60.9%; P=.03).

Overall engagement over the immediate postoperative period was similar between the messaging arms (Table 2). Using automated bidirectional SMS text messaging, the overall response rate was 45.3% (29/64) versus 49.2% (31/63) using a hyperlink to a mobile survey. For those patients using opioids in the past 4 days, prolonged future engagement was higher in the mobile survey arm at day 7 (18/19, 94.7% vs 9/15, 60%). Among nonresponders, the majority of patient drop-off occurred at day 4. Patients with more invasive or hip procedures showed a trend toward being responsive at day 4 as compared to those not responding (Table 3).

Table 1. Patient demographics.

Demographic	Bidirectional SMS text messaging (N=64)	Mobile survey (N=63)	P value
Sex, n (%)			.72
Female	26 (40.6)	28 (44.4)	
Male	38 (59.4)	35 (55.6)	
Age (years), mean (SD)	37.8 (14.2)	39.2 (13.7)	.58
Comorbidities, n (%)			.49
0	29 (45.3)	22 (34.9)	
1-2	13 (20.3)	16 (25.4)	
>2	22 (34.4)	25 (39.7)	
Type of surgery, n (%)			.02
Knee	40 (62.5)	50 (83.3)	
Shoulder	19 (29.7)	6 (10)	
Hip	5 (7.8)	4 (6.7)	
Elbow	0 (0)	2 (3.2)	
Invasive procedure (open or nonlaparoscopic), n (%)	39 (60.9)	26 (41.3)	.03
Quantity of opioid tablets[a] prescribed, n (%)			.09
<11	15 (23.4)	23 (36.5)	
11-20	25 (39.1)	28 (44.4)	
21-30	19 (29.7)	8 (12.7)	
>30	5 (7.8)	4 (6.4)	

[a]Opioid tablet in 5 mg oxycodone equivalents.

Table 2. Patient-reported outcomes and response rates.

Day	Conversational messaging					Mobile survey				
	Total, n	No response, n	Completed[a], n	Continue on[b], n	Response rate, %	Total, n	No response, n	Completed[a], n	Continue on[b], n	Response rate, %
Day 4	64	28	21	15	56.3	63	30	14	19	52.4
Day 7	16	7	5	4	56.3	19	1	10	8	94.7
Day 14	4	0	3	1	100	8	1	5	2	87.5
Day 21	1	1	0	0	0	2	0	1	1	100
Day 28	0	N/A[c]	N/A	N/A	N/A	1	0	1	0	100
Final response	N/A	34	30	N/A	46.9	N/A	32	31	N/A	49.2

[a]No longer taking opioids or no longer planning to take them.

[b]Participants indicating continued or planned use of opioids, who were thus sent additional surveys on subsequent days.

[c]N/A: not applicable.

Table 3. Response at day 4 by demographics.

Demographic	No response (N=58)	Responded day 4 (N=69)	P value
Arm, n (%)			.72
Bidirectional messaging	28 (48.3)	36 (52.2)	
Mobile survey	30 (51.7)	33 (47.8)	
Sex, n (%)			.37
Female	22 (37.9)	32 (46.4)	
Male	36 (61.1)	37 (53.6)	
Age (years), mean (SD)	38.1 (13.8)	38.9 (14)	.76
Comorbidities, n (%)			.74
0	23 (39.7)	28 (40.6)	
1-2	15 (25.9)	14 (20.3)	
>2	20 (34.5)	27 (39.1)	
Type of surgery, n (%)			.09
Knee	44 (75.9)	46 (66.7)	
Shoulder	12 (20.7)	13 (18.8)	
Hip	1 (1.7)	9 (13)	
Elbow	1 (1.7)	1 (1.5)	
Invasive procedure (open or nonlaparoscopic), n (%)	24 (41.4)	41 (59.4)	.05
Quantity of opioid tablets[a] prescribed, n (%)			.69
<11	20 (34.5)	18 (26.1)	
11-20	24 (41.4)	29 (42)	
21-30	11 (19)	16 (23.2)	
>30	3 (5.2)	6 (8.7)	

[a]Opioid tablet in 5 mg oxycodone equivalents.

Discussion

Principal Findings

This study has two key findings within the context of mobile patient engagement and data capture. First, immediate postoperative patient engagement and research consent are feasible using digital methods. Second, there were no significant differences in the overall response rates between the modalities of bidirectional SMS text messaging and mobile surveys. We deployed a direct-to-patient approach in obtaining mobile consent to prospectively capture PROs for postoperative pain and pain management. Prior studies aiming to understand patients' pain and use of prescription opioids following surgeries have been limited by retrospective design, telephone or paper surveys, and recall bias [19,22,27]. We worked collaboratively within the health system to develop an approach that offloads clinical providers from obtaining written consent during preoperative visits and reaches out to patients following their procedure through SMS text messaging. This mobile consent process may be further studied to decrease in-person time, allow for researchers to link important research and other patient information, and offer patients the ability to opt out.

Second, the method and approach used to capture patient-reported data may not significantly impact initial response rates and overall completion rates. The ways in which patients use digital technology and mobile apps continue to change [28]. This study begins to analyze direct-to-patient methods that capture patient-reported data [29]. These early results indicate mobile methods can be used to engage postoperative patients and may provide a scalable approach for engaging larger patient populations. Traditional paper-based surveys rely on mail services and may introduce time delays and recall bias, whereas digital methods allow patients to respond in the moment [30]. Though not statistically significant, we found that patients undergoing more invasive procedures (ie, open surgeries and nonlaparoscopic surgeries) or any procedure on the hip were more likely to respond and remain engaged. This suggests an opportunity to provide tailored content and messaging as more procedures become outpatient and more recovery is based in the home. Ultimately, we describe similar and consistent completion rates as the weeks passed following patients' surgeries.

Limitations

This study has limitations. First, the study and data collection were performed at a single academic medical center and thus the study is not generalizable. Second, selection bias may be

present as patients needed to have a mobile device and opt to answer questions through either the mobile survey or the bidirectional SMS text messaging system. Those patients receiving a mobile link to the survey must have a smartphone and may represent a select population. Third, patients were undergoing elective outpatient procedures and may represent a population that is not generalizable to all procedures. This study compares two techniques that have been rapidly adopted; it does not compare these directly to paper methods, which have more traditionally been used.

Conclusion

This study demonstrates the early feasibility of PRO capture using two methods of patient engagement following orthopedic surgery. The findings suggest no major differences between bidirectional SMS text messaging and mobile surveys to collect PRO data. In an increasingly digital era, clinicians and researchers may employ digital surveys as a tool for rapid collection of PROs, which can inform research and a learning health system. Future studies will need to investigate larger scale programs and generalizability outside of surgical settings and for broader populations.

Authors' Contributions

AKA and MKD researched the literature and conceived the study. ZSA, JH, FS, RX, DAR, and ES were involved in protocol development, gaining ethical approval, patient recruitment, and data analysis. AKA wrote the first draft of the manuscript. All authors reviewed and edited the manuscript and approved the final version of the manuscript. MKD, ZSA, and AKA are the guarantors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT eHEALTH Checklist (V 1.6.1).
[PDF File (Adobe PDF File), 690 KB - formative_v6i3e31894_app1.pdf]

References

1. Olsen L, Aisner D, McGinnis JM. The Learning Healthcare System: Workshop Summary. In: Institute of Medicine (US) Roundtable on Evidence-Based Medicine. Washington (DC): National Academies Press (US); 2007 Presented at: The Learning Health System Workshop; 2007; Washington, DC. [doi: 10.17226/11903]
2. About Learning Health Systems. Agency for Healthcare Research and Quality. URL: http://www.ahrq.gov/learning-health-systems/about.html [accessed 2019-04-09]
3. Free C, Knight R, Robertson S, Whittaker R, Edwards P, Zhou W, et al. Smoking cessation support delivered via mobile phone text messaging (txt2stop): a single-blind, randomised trial. Lancet 2011 Jul 2;378(9785):49-55 [FREE Full text] [doi: 10.1016/S0140-6736(11)60701-0] [Medline: 21722952]
4. Basch E. Patient-Reported Outcomes - Harnessing Patients' Voices to Improve Clinical Care. N Engl J Med 2017 Jan 12;376(2):105-108. [doi: 10.1056/NEJMp1611252] [Medline: 28076708]
5. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment. JAMA 2017 Dec 11;318(2):197-198 [FREE Full text] [doi: 10.1001/jama.2017.7156] [Medline: 28586821]
6. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom Monitoring With Patient-Reported Outcomes During Routine Cancer Treatment: A Randomized Controlled Trial. J Clin Oncol 2016 Feb 20;34(6):557-565. [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
7. Demographics of Mobile Device Ownership and Adoption in the United States. Pew Research Center. URL: https://www.pewinternet.org/fact-sheet/mobile/ [accessed 2019-08-27]
8. Americans and Text Messaging. Pew Research Center. URL: http://www.pewinternet.org/2011/09/19/americans-and-text-messaging/ [accessed 2019-01-28]
9. Bernstein DN, Fear K, Mesfin A, Hammert WC, Mitten DJ, Rubery PT, et al. Patient-reported outcomes use during orthopaedic surgery clinic visits improves the patient experience. Musculoskeletal Care 2019 Mar;17(1):120-125. [doi: 10.1002/msc.1379] [Medline: 30615249]

10. Calvert MJ, O'Connor DJ, Basch EM. Harnessing the patient voice in real-world evidence: the essential role of patient-reported outcomes. Nat Rev Drug Discov 2019 Sep;18(10):731-732. [doi: 10.1038/d41573-019-00088-7] [Medline: 31570837]

11. NeCamp T, Sen S, Frank E, Walton MA, Ionides EL, Fang Y, et al. Assessing Real-Time Moderation for Developing Adaptive Mobile Health Interventions for Medical Interns: Micro-Randomized Trial. J Med Internet Res 2020 Mar 31;22(3):e15033 [FREE Full text] [doi: 10.2196/15033] [Medline: 32229469]

12. Cao J, Truong AL, Banu S, Shah AA, Sabharwal A, Moukaddam N. Tracking and Predicting Depressive Symptoms of Adolescents Using Smartphone-Based Self-Reports, Parental Evaluations, and Passive Phone Sensor Data: Development and Usability Study. JMIR Ment Health 2020 Jan 24;7(1):e14045 [FREE Full text] [doi: 10.2196/14045] [Medline: 32012072]

13. Rohani DA, Faurholt-Jepsen M, Kessing LV, Bardram JE. Correlations Between Objective Behavioral Features Collected From Mobile and Wearable Devices and Depressive Mood Symptoms in Patients With Affective Disorders: Systematic Review. JMIR mHealth uHealth 2018 Aug 13;6(8):e165 [FREE Full text] [doi: 10.2196/mhealth.9691] [Medline: 30104184]

14. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile Telephone Text Messaging for Medication Adherence in Chronic Disease: A Meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]

15. van Onzenoort HAW, Verberk WJ, Kessels AGH, Kroon AA, Neef C, van der Kuy PM, et al. Assessing medication adherence simultaneously by electronic monitoring and pill count in patients with mild-to-moderate hypertension. Am J Hypertens 2010 Feb;23(2):149-154. [doi: 10.1038/ajh.2009.207] [Medline: 19927136]

16. Agarwal A, Patel M. Prescribing Behavior Change: Opportunities and Challenges for Clinicians to Embrace Digital and Mobile Health. JMIR mHealth uHealth 2020 Aug 04;8(8):e17281 [FREE Full text] [doi: 10.2196/17281] [Medline: 32749997]

17. Abdolkhani R, Gray K, Borda A, DeSouza R. Patient-generated health data management and quality challenges in remote patient monitoring. JAMIA Open 2019 Dec;2(4):471-478 [FREE Full text] [doi: 10.1093/jamiaopen/ooz036] [Medline: 32025644]

18. Asch DA, Muller RW, Volpp KG. Automated hovering in health care--watching over the 5000 hours. N Engl J Med 2012 Jul 05;367(1):1-3. [doi: 10.1056/NEJMp1203869] [Medline: 22716935]

19. Waljee JF, Zhong L, Hou H, Sears E, Brummett C, Chung KC. The Use of Opioid Analgesics following Common Upper Extremity Surgical Procedures: A National, Population-Based Study. Plast Reconstr Surg 2016 Feb;137(2):355e-364e [FREE Full text] [doi: 10.1097/01.prs.0000475788.52446.7b] [Medline: 26818326]

20. Thiels CA, Ubl DS, Yost KJ, Dowdy SC, Mabry TM, Gazelka HM, et al. Results of a Prospective, Multicenter Initiative Aimed at Developing Opioid-prescribing Guidelines After Surgery. Ann Surg 2018 Sep;268(3):457-468. [doi: 10.1097/SLA.0000000000002919] [Medline: 30004924]

21. Thiels CA, Anderson SS, Ubl DS, Hanson KT, Bergquist WJ, Gray RJ, et al. Wide Variation and Overprescription of Opioids After Elective Surgery. Ann Surg 2017 Oct;266(4):564-573. [doi: 10.1097/SLA.0000000000002365] [Medline: 28697049]

22. Bicket MC, Long JJ, Pronovost PJ, Alexander GC, Wu CL. Prescription Opioid Analgesics Commonly Unused After Surgery: A Systematic Review. JAMA Surg 2017 Nov 01;152(11):1066-1071 [FREE Full text] [doi: 10.1001/jamasurg.2017.0831] [Medline: 28768328]

23. Schwartzenberger J, Presson A, Lyle A, O'Farrell A, Tyser AR. Remote Collection of Patient-Reported Outcomes Following Outpatient Hand Surgery: A Randomized Trial of Telephone, Mail, and E-Mail. J Hand Surg Am 2017 Sep;42(9):693-699. [doi: 10.1016/j.jhsa.2017.05.002] [Medline: 28600107]

24. Agarwal AK, Lee D, Ali Z, Sennett B, Xiong R, Hemmons J, et al. Patient-Reported Opioid Consumption and Pain Intensity After Common Orthopedic and Urologic Surgical Procedures With Use of an Automated Text Messaging System. JAMA Netw Open 2021 Mar 01;4(3):e213243 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.3243] [Medline: 33764425]

25. Agarwal AK, Ali ZS, Sennett B, Xiong R, Hemmons J, Spencer E, et al. An Automated Text Messaging Program to Inform Postoperative Opioid Prescribing. NEJM Catalyst 2021 Feb 17;2(3):A. [doi: 10.1056/cat.20.0440]

26. Calculating Total Daily Dose of Opioids For Safer Dosage. Centers for Disease Control and Prevention. 2021 Feb 17. URL: https://www.cdc.gov/drugoverdose/pdf/calculating_total_daily_dose-a.pdf [accessed 2022-03-11]

27. Cramer JD, Mehta V, Chi DH, Weinreich HM. Minimizing Excess Opioid Prescribing for Acute Postoperative Pain. JAMA Otolaryngol Head Neck Surg 2020 Mar 01;146(3):228-230. [doi: 10.1001/jamaoto.2019.4209] [Medline: 31917416]

28. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]

29. Morgan AU, Balachandran M, Do D. Remote Monitoring of Patients with Covid-19: Design, implementation, and outcomes of the first 3,000 patients in COVID Watch. NEJM Catalyst 2020:A [FREE Full text]

30. Howard R, Fry B, Gunaseelan V, Lee J, Waljee J, Brummett C, et al. Association of Opioid Prescribing With Opioid Consumption After Surgery in Michigan. JAMA Surg 2019 Jan 01;154(1):e184234 [FREE Full text] [doi: 10.1001/jamasurg.2018.4234] [Medline: 30422239]

XSL•FO

RenderX

Abbreviations

PRO: patient-reported outcome

XSL•FO
RenderX

<u>Original Paper</u>

An Integrated, Multimodal, Digital Health Solution for Chronic Obstructive Pulmonary Disease: Prospective Observational Pilot Study

Brian D Gelbman[1*], MD; Carol R Reed[2*], MD

[1]Division of Pulmonary and Critical Care Medicine, Weill Cornell Medical Center, New York, NY, United States

[2]Wellinks (Convexity Scientific, Inc), New Haven, CT, United States

[*]all authors contributed equally

Corresponding Author:
Brian D Gelbman, MD
Division of Pulmonary and Critical Care Medicine
Weill Cornell Medical Center
635 Madison Avenue, Suite 1101
New York, NY, 10022
United States
Phone: 1 212 628 6611
Email: Brg9001@med.cornell.edu

Abstract

Background: Chronic obstructive pulmonary disease (COPD) affects millions of Americans and has a high economic impact partially due to frequent emergency room visits and hospitalizations. Advances in digital health have made it possible to collect data remotely from multiple devices to assist in managing chronic diseases such as COPD.

Objective: In this pilot study, we evaluated the ability of patients with COPD to use the Wellinks mHealth platform to collect information from multiple modalities important to the management of COPD. We also assessed patient satisfaction and engagement with the platform.

Methods: A single-site, observational, prospective pilot study (N=19) was conducted using the Wellinks platform in adults with COPD. All patients were aged over 30 years at screening, owned an iPhone, and were currently undergoing a treatment regimen that included nebulized therapy. Enrolled patients received a study kit consisting of the Flyp nebulizer, Smart One spirometer, the Nonin pulse oximeter, plus the Wellinks mHealth app, and training for all devices. For 8 weeks, participants were to enter daily symptoms and medication use manually; spirometry, nebulizer, and pulse oximeter data were automatically recorded. Data were sent to the attending physician in a monthly report. Patient satisfaction was measured via a 5-point scale and the Net Promoter Score (NPS) captured in interviews at the end of the observation period.

Results: Average age of the patients was 79.6 (range 65-95) years. Participants (10 female; 9 male) had an average FEV_1% (forced expiratory volume in 1 second as % of predicted for the patient) of 56.2% of predicted (range 23%-113%) and FEV_1/forced vital capacity of 65%. COPD severity, as assessed by the Global Initiative for Chronic Obstructive Lung Disease (GOLD) classification, was mild in 2 patients, moderate in 6, and severe/very severe in 11; 9 patients were on home oxygen. During this 8-week study, average use of the spirometer was 2.5 times/week, and the pulse oximeter 4.2 times/week. Medication use was manually documented 9.0 times/week, nebulizer use 1.9 times/week, and symptoms recorded 1.2 times/week on average. The correlation coefficients of home to office measurements for peak flow and FEV_1 were high (r=0.94 and 0.96, respectively). Patients found the app valuable (13/16, 81%) and easy to use (15/16, 94%). The NPS was 59.

Conclusions: This study demonstrates that our cohort of patients with COPD engaged with the Wellinks mHealth platform avidly and consistently over the 8-week period, and that patient satisfaction was high, as indicated by the satisfaction survey and the NPS of 59. In this small, selected sample, patients were both willing to use the technology and capable of doing so successfully regardless of disease severity, age, or gender. The Wellinks mHealth platform was considered useful and valuable by patients, and can assist clinicians in improved, timely decision making for better COPD management.

Introduction

Chronic obstructive pulmonary disease (COPD) is a respiratory disorder characterized by persistent symptoms such as shortness of breath, coughing, excess mucus production, and irreversible expiratory airflow obstruction. It primarily affects people aged 65 and over and is often a result of exposure to risk factors such as tobacco smoke or air pollutants [1]. In 2018 in the United States, 16.4 million people reported having a diagnosis of COPD, chronic bronchitis, or emphysema, according to the Centers for Disease Control and Prevention's National Health Interview Survey [2] and COPD was the fourth leading cause of death in the United States [3].

COPD has a high financial impact. COPD-related costs in the United States were projected to be US $49.0 billion in 2020, including direct health care expenditures and indirect morbidity and mortality costs [4]. COPD exacerbations, during which symptoms of coughing and shortness of breath acutely worsen, are a significant driver of these expenditures, due to emergency room visits, hospital admissions, relapses, and readmissions [5]. COPD has a 30-day readmission rate of approximately 20% and is a target condition in the Medicare Hospital Readmission Reduction Program [6]. Frequent COPD exacerbations can lead to a decline in lung function and quality of life and may significantly affect a patient's prognosis [7,8].

The current standard of care for COPD includes medications, smoking cessation, oxygen supplementation (when indicated), and pulmonary rehabilitation [9]. At each office visit, patient biometric data are collected (eg, pulse oximetry and spirometry) and the COPD care plan is communicated. Between office visits, however, it is difficult for the physician to gather insights into patient behavior and adherence to the care plan; significant clinical deterioration or noncompliance can go undetected. Furthermore, many patients with COPD have limited mobility due to the debilitating nature of their lung disease, making office visits challenging or not possible at all. Currently there are few comprehensive and reliable tracking solutions to assess patient adherence to the care plan or the impact of medication and other interventions on a more frequent basis [8,10,11].

Advances in digital health have now made it possible to share medical information in real time, thus enabling more immediate management of chronic diseases. Remote patient monitoring solutions have been successfully developed in other chronic diseases, such as congestive heart failure and diabetes [12,13]. During the COVID-19 epidemic, patients of all demographic groups quickly adopted some form of technology to access medical care [14,15]. Monitoring modalities are now being evaluated for factors such as ease of integration into existing clinician workstreams, correlation with traditional diagnostic measures and treatment modalities, and overall patient acceptance [16].

For patients with COPD, effective remote monitoring should include information on medication adherence and symptoms, plus biometric data including spirometry and pulse oxygen [17]. However, the ability to monitor data streams from different digital sources has presented some challenges, such as accuracy, measuring adherence, and ease of use for both the patient and the clinician, including incorporation into existing electronic medical records [8,11,17].

In this pilot study, we evaluated the Wellinks mHealth platform in a small cohort of patients with COPD. Our objective was to, within a trial setting, determine the feasibility of the platform to track information important to the management of COPD, collected via Bluetooth from multiple modalities including spirometry and pulse oximetry, medication adherence, and patient input of symptoms and medication use. This small-scale study was performed to obtain feedback on patient satisfaction and engagement with the platform, and to inform modifications needed for larger clinical studies.

Methods

Study Design

This prospective observational pilot study of the Wellinks mHealth Platform was conducted from January to May 2021 at a single outpatient pulmonary practice. A sample size of 20 patients was planned. Monitoring data were collected over an 8-week-per-study-patient observation period. Informed consent was obtained from each participant, and the study was conducted according to the principles stated in the Declaration of Helsinki (2013). Patients were deidentified of protected health information according to the Safe Harbor method of the Health Insurance Portability and Accountability Act (HIPAA). Each patient was assigned a unique study ID number and the principal investigator maintained sole access to reidentification codes.

Ethical Approval

This study was approved by the Institutional Review Board (IRB) (IRB#:20-WELL-101) with respect to the scientific content and ethical treatment of human research participants.

Recruitment

Participants were recruited by the principal investigator within the clinical setting. Inclusion criteria specified male or female patients with COPD over 30 years of age with English language literacy who were prescribed a treatment regimen that included nebulized therapy. All patients had access to an iPhone running iOS version 13.4 or later.

Exclusion criteria were acute or chronic conditions that might interfere with patients' ability to participate in the study, or comorbidities that might interfere with data collection and interpretation (eg, renal, cardiac, hepatic, central nervous system, or psychiatric conditions).

Participation in this study was completely voluntary. Participants were permitted to keep the items in the study kit. No additional financial inducements were offered, and no patient recruitment materials were used.

XSL•FO

RenderX

Patient Use of the Intervention

After informed consent was obtained, participants received the study kit, which consisted of 3 devices: the Flyp nebulizer, MIR Smart One spirometer, and Nonin pulse oximeter, plus instructions for use (Figure 1). In addition, participants received instructions for downloading the Wellinks mHealth iOS app. The physician or another member of the research team explained the function of all devices and the app. Participants were considered enrolled once the app was downloaded and the participant had logged in.

Figure 1. Components of the Wellinks mHealth Kit, including the Flyp nebulizer, MIR Smart One spirometer, and Nonin pulse oximeter, are pictured here from left to right. Patients used their own phones to download the Wellinks mHealth App.

Patients were asked to perform forced expiratory volume in 1 second (FEV_1) and pulse oximetry measurements (peripheral oxygen saturation or SpO_2) at least once weekly; these data were automatically recorded in the app via Bluetooth. Patients' use of the Flyp nebulizer was also automatically recorded; most patients were prescribed nebulizer treatments as rescue therapy only.

Daily medication usage and symptoms were entered into the app manually by the patient (Figure 2). The app was prepopulated with a customized list of medications prescribed to each patient for COPD, including dosing intervals; patients were able to check off when the dose of each prescribed medication was taken. Patients could also enter the use of additional medications, such as rescue inhalers or nebulizer treatments. A list of symptoms (mucus production, shortness of breath, chest tightness, wheezing, coughing, low energy, and trouble sleeping) allowed patients to either check the symptom boxes or to check a "no symptoms" box if they were not having symptoms. Patients were not prompted by the app or the clinical research team to use the devices or to enter information into the app.

Figure 2. The app displays content that guides the patient while recording pulse oximetry, or inputting medications taken and symptoms experienced from a list. The medication list was prepopulated with each patient's prescribed medications.
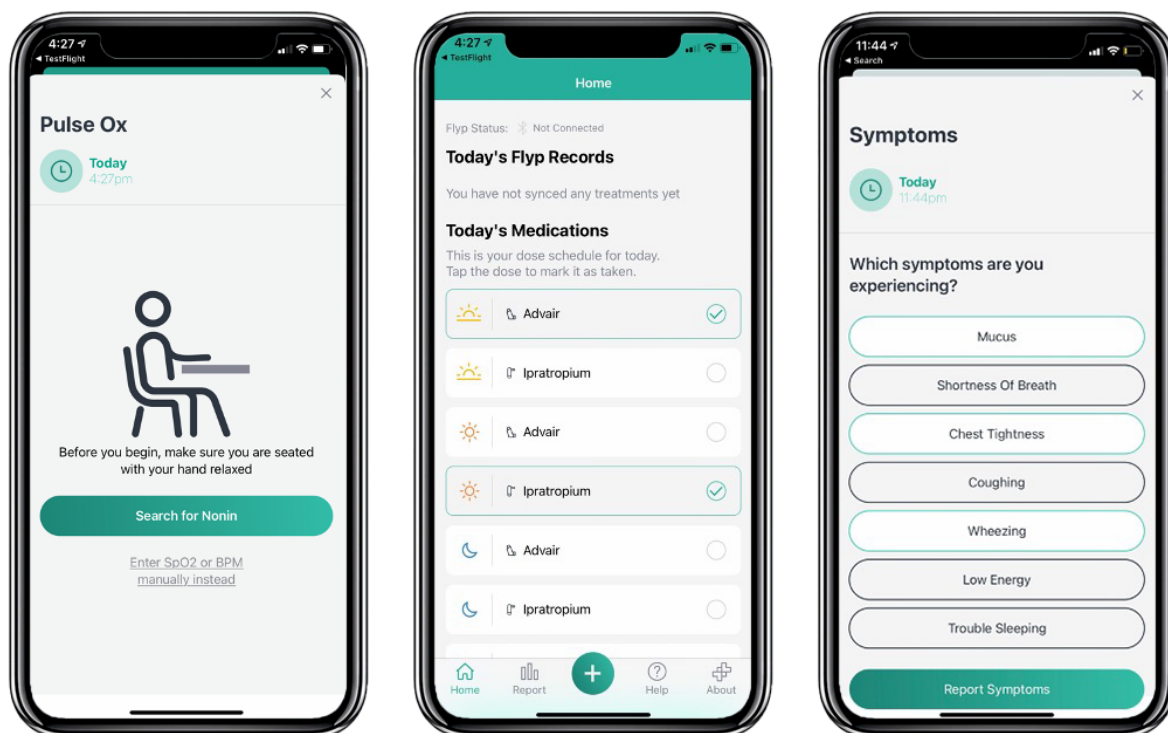
The Wellinks app collects and stores data related to treatment and respiratory status of the patient on a secure, HIPAA-compliant, cloud-hosted database. Monthly reports containing a summary of the collected medication adherence data, FEV_1 and SpO_2 measurements, and symptoms or notes were sent to the clinician via secure email. Any subsequent visits or treatment were to be consistent with current standard of care (Figure 3).

Figure 3. The Wellinks mHealth platform recorded spirometry, pulse oximetry, and nebulizer use automatically by bluetooth. The patient manually input medications taken and symptoms experienced (if any) using the app. Data were stored securely and compiled into a summary Portable Document Format (PDF) report for the clinician monthly. API: application programming interface, HIPAA: Health Insurance Portability and Accountability Act.
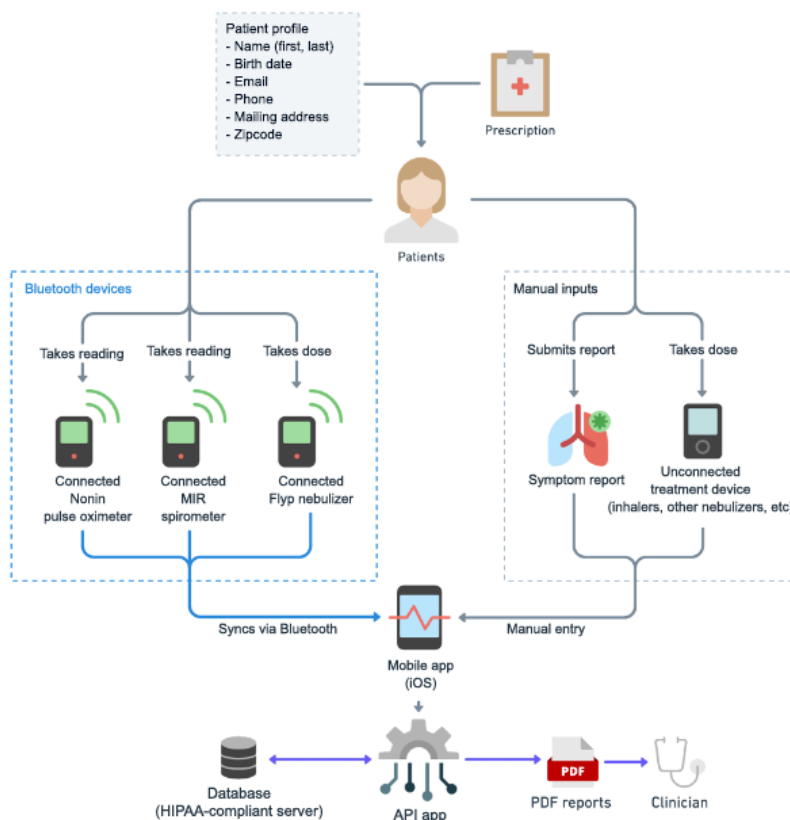

Patient Engagement and Satisfaction

Patient engagement was evaluated by assessing how often patients used the various features of the Wellinks mHealth platform over the 8-week study period. For spirometry and pulse oximetry, the baseline case was 1 use/week. Use of the medication and symptoms recording feature was left to the patient's discretion.

At the end of the observation period, patients' overall attitudes about and perception of the Wellinks mHealth solution were assessed using a 9-question 5-point Likert scale (responses ranged from 1=strongly disagree to 5=strongly agree). In addition, patients had the opportunity to provide qualitative feedback to the interviewer regarding their experience and the intervention.

The Net Promoter Score (NPS) was used to gauge patient satisfaction overall with the Wellinks mHealth solution. The NPS is derived from the answer to just 1 question: "On a scale of 0-10, how likely is it that you would recommend using the Wellinks app to friends, family members, or associates also living with COPD?". The score is determined by subtracting the percentage of "Detractors" (those who gave a score of ≤6) from the "Promoters" (those who gave a score of 9 or 10). The "Passives" (those who gave a score of 7 or 8) are thought to be satisfied with the product, but not to the point of recommending the product or "promoting" it to others [18,19]. With the methodology used here, the total score (tallied from all responses) will fall between –100 and 100, with –100 being the worst outcome and 100 being the best. This survey was administered to the patient by video call or telephone by a member of the Wellinks research team.

The Wellinks mHealth platform generated a physician report monthly (Figure 4). Each report was customized for each patient, and provided the following summary of that patient's activity in the preceding month:

- Scheduled medications (name and dosage): adherence percentage and medication most commonly missed;
- As-needed (pro re nata [PRN]) medications (name and dosage): how often and how much taken;
- SpO_2% over time;
- Pulse (beats/minute) over time;
- Spirometry (peak flow and FEV_1) over time; and
- Symptoms: list of symptoms experienced and when entered (date and time).

Figure 4. The physician report supplied monthly summarizes prescribed scheduled medications, PRN medications and adherence, pulse oximetry and spirometry readings over time, and symptoms. PRN: pro re nata.
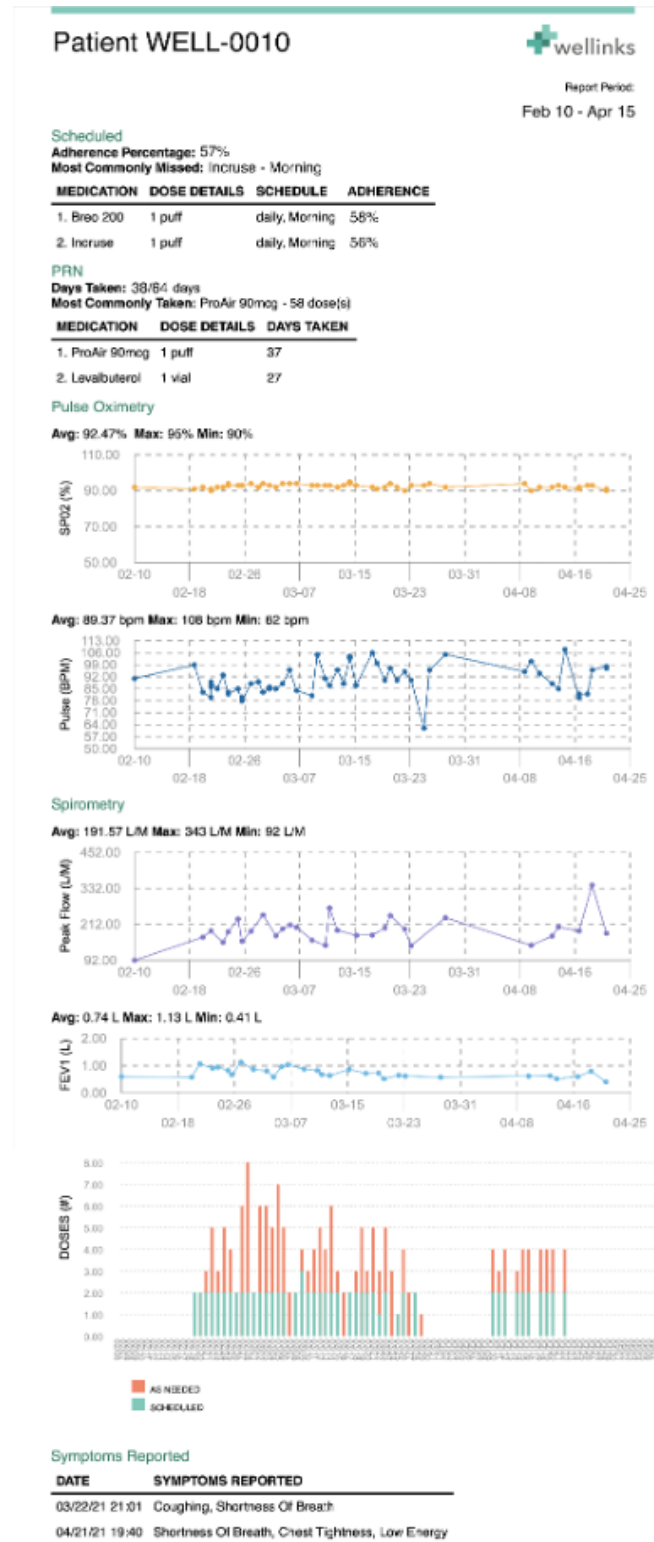
Adverse Events

Adverse events occurring during the study period were documented by the investigator; these were not reported to the device manufacturer unless they were considered serious and related to the device. Serious adverse events were to be reported using an unanticipated device effects form and were to be reported to the sponsor and to the IRB.

Analysis

Data were summarized with means and ranges for quantitative variables with normal distribution (eg, age, spirometry values). The Pearson correlation coefficients were calculated to compare concordance between home and office spirometry; mean values from the Smart One spirometer were compared with FEV_1 values collected at the closest patient visit date (office spirometry was performed on a Vyaire Vmax Encore).

Correlations between office and home pulse oximetry were not conducted because of the narrow range of values. The *t* tests were performed to compare differences between subgroups using Microsoft Excel. *P* values are given for subgroup analyses, with a threshold of .05 considered statistically significant.

Results

Demographics

A total of 19 patients were enrolled in the study. The average age was 79.6 years, and the patient group was almost equally divided by gender (10 female, 9 male). Patients' $FEV_1\%$ (forced expiratory volume in 1 second as % of predicted for the patient) at study entry averaged 56.2% of predicted (range 23%-113%), and FEV_1/forced vital capacity averaged 65%. The majority of patients (n=10, 53%) fell into the severe disease category, as indicated by the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria [9]. A total of 6 patients (32%) were in the GOLD 2, or moderate category, and 2 (11%) were in the mild category; 1 patient (5%) was categorized as GOLD 4, or very severe. This distribution is not surprising as patients were required to be using nebulizers, which are not typically prescribed for mild disease. A total of 5 patients had a caregiver assist them with many of the activities, including equipment use and symptom/medication input (Table 1).

Table 1. Demographics and baseline characteristics of the study population.

Demographic characteristic	Value (N=19)
Age in years, mean (range)	79.6 (65-95)
Gender, n (%)	
Male	9 (47)
Female	10 (53)
GOLD[a] category, n (%)[b]	
GOLD 1: mild (FEV_1[c] ≥80% predicted)	2 (11)
GOLD 2: moderate (FEV_1 ≥50-79%)	6 (32)
GOLD 3: severe (FEV_1 ≥30-49%)	10 (53)
GOLD 4: very severe (FEV_1 <30%)	1 (5)
$FEV_1\%$[d], mean (range)	56.2 (23-113)
FEV_1/forced vital capacity, mean	65.3
Home oxygen use, n (%)	9 (47)
Pack years, mean (range)	51.6 (10-100)
Independent versus assisted use of Wellinks mHealth solution, n (%)	
Independent use	14 (74)
Assisted use (with help from a caregiver or home health aide)	5 (26)

[a]GOLD: Global Initiative for Chronic Obstructive Lung Disease.

[b]All patients with FEV_1/forced vital capacity <0.70.

[c]FEV_1: forced expiratory volume in 1 second.

[d]$FEV_1\%$: forced expiratory volume in 1 second as % of predicted for the patient.

Patient Use of the Intervention

Participants used the pulse oximeter 4.2 times/week and the spirometer 2.5 times per week on average (Table 2). All 19 patients achieved the weekly goal of at least one pulse oximetry reading per week over 8 weeks, and 15 patients met or exceeded the goal of 1 spirometry reading per week. On average, patients' medication use was entered into the app 9 times/week, and symptoms entered 1.2 times/week.

Differences in participation did not significantly differ by COPD severity, age, or gender (Table 3).

There was a strong correlation between the FEV_1 (r=0.96) and peak flow (r=0.94) measurements recorded by the spirometer compared with the measurements recorded by the physician during the office visit closest in time to the at-home collected information (Figure 5).

Table 2. Study participant interaction with the mobile health platform.

Data collected	Mean number of app entries/week (range of number of app entries/week)
FEV_1[a] by spirometer	2.5 (1-7)
Blood oxygenation by pulse oximeter	4.2 (1-12)
Medication use	9.0 (1-25.1)
Nebulizer use	1.9 (0-11.9)
Symptoms	1.2 (0-5.6)

[a]FEV_1: forced expiratory volume in 1 second.

Table 3. Participation by gender, age, and COPD[a] severity.

Demographic characteristic	n	Spirometry[b]	Pulse oximetry[b]	Medications[c]	Symptoms[c]
Gender					
Male	9	19.3	35.2	10.3	6.5
Female	10	21.1	32.6	21.1	12.2
P value		.81	.84	.37	.22
Age					
<80	11	25.4	41.6	22.8	11.6
≥80	8	13.0	23.5	5.3	5.8
P value		.09	.15	.14	.22
COPD severity					
Mild + moderate	8	19.8	28.5	17.1	5.5
Severe + very severe	11	20.5	37.9	14.1	11.9
P value		.93	.46	.81	.17

[a]COPD: chronic obstructive pulmonary disease.

[b]Data are presented as mean number of uses/week.

[c]Data are presented as mean number of times reported.

Figure 5. The correlation of home versus office assessments of (A) peak flow and (B) FEV_1. FEV_1: forced expiratory volume in 1 second.
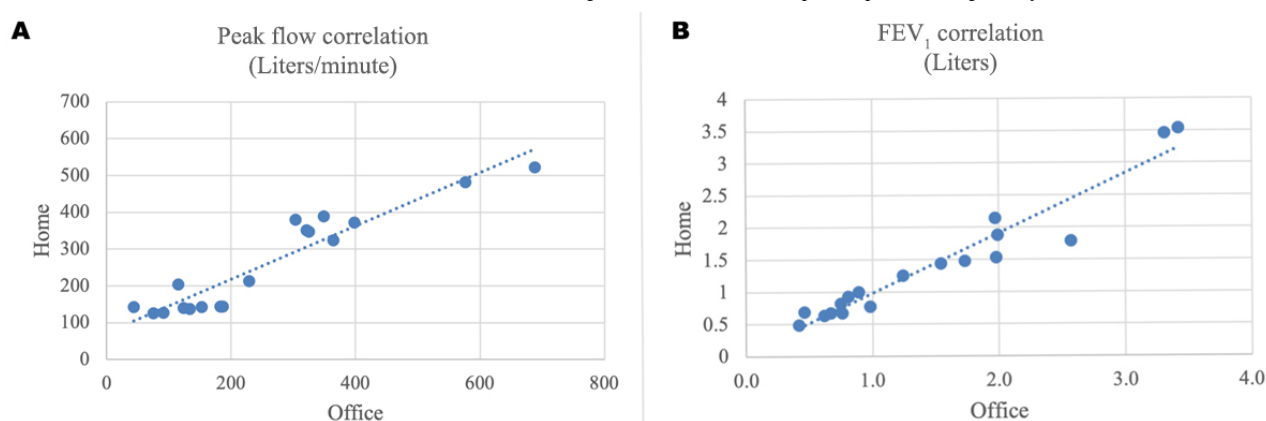
Patient Engagement and Satisfaction

The average number of recordings for spirometry, oximetry, and medication usage declined over the 8-week interval by varying degrees (Figure 6). Medication use entries fell from 7.8 times/week to 3.7 times/week (a reduction of 52.3%), oximetry recordings from 5.5 times/week to 2.5 times/week (a reduction of 54.2%), and spirometry recordings from 3.4 times/week to 1.8 times/week (a reduction of 45.4%).

Figure 6. Engagement, measured in app uses or recordings/week, fell over the course of the 8-week study, but remained above the baseline requested of 1 use/week for spirometry and oximetry recordings.
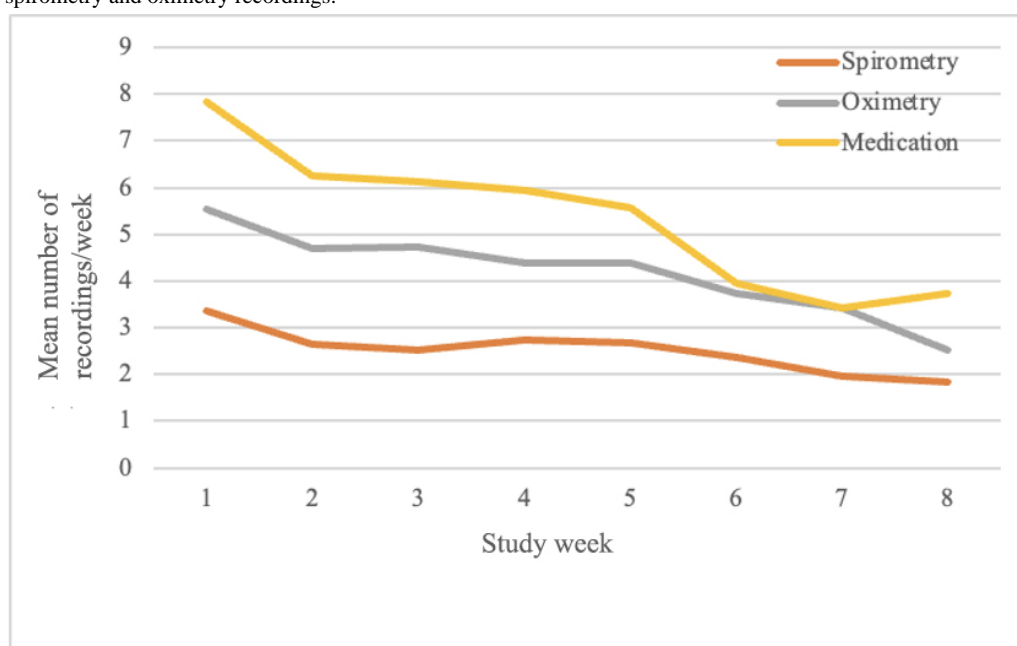
The satisfaction survey administered is presented as Multimedia Appendix 1. A total of 16 patients were available to take this survey. Most patients (15/16, 94%) either agreed or strongly agreed that the Wellinks app was easy to use and valuable (13/16, 81%). Most (15/16, 94%) found that it was valuable to be able to take spirometry and pulse oximetry measurements at home. The symptom logging function was found to be moderately valuable (11/16, 69%) as was the medication schedule (10/16, 63%). Patients expressed interest (12/16, 75%) in adding a physician or care team messaging component to the app. In general, the patients did not feel that the app enhanced their knowledge of COPD or the connection with their doctor. The NPS generated by the patients in this study was 59.

Adverse Events

There were no issues reported by either patients or the principal investigator related to the safety of the patients or the performance of the devices.

Discussion

Principal Findings

COPD is a chronic condition where the potential for rapid deterioration due to exacerbations is an ever-present danger [6,7,9]. For this reason, it can be a very difficult condition to manage as adjustments in care are dependent on the patient recognizing and reporting changes from baseline. While various devices exist to enable monitoring of specific individual parameters, an app that allows patients and physicians to track multiple modalities simultaneously, and remotely, would result in the most complete clinical picture available from the patient's daily life.

In this pilot study, we observed that study participants were both willing to use the Wellinks mHealth solution and able to do so effectively, performing both spirometry and oximetry measurements far more often than requested, and logging

medication use and symptoms frequently without being prompted. Patients engaged with the Wellinks platform regardless of disease severity, gender, or age. Notably, the correlations of peak flow and FEV_1 measurements taken by the patient at home and those performed in the office were very high. This has been a technical failure in prior studies, where use of the spirometer in an outpatient setting and without the assistance of a health care practitioner has resulted in inaccurate results [20,21].

Age may be a concern in the adoption of an mHealth system in patients with COPD, as COPD is largely a disease of the elderly who may have a lower digital literacy level, less access to smart devices, and overall, a lower comfort level with technology. A 2018 study of 638,330 Medicare beneficiaries found that 40.9% lacked a smartphone with a wireless data plan, and 26.3% lacked any type of digital access (either a smartphone or a home desktop or laptop computer with a high-speed internet connection) [22].

However, the COVID-19 pandemic has changed the way the elderly view technology. In a recent study conducted by the American Association for Retired People (AARP), 45% of individuals over 60 years viewed technology far more positively than they did prior to COVID. Among those 70 and over, 73% reported having high-speed internet access at home. Ownership of smart technology, including smartphones, continues to increase among those over age 70, and accessing health care services and health information is among the top 10 activities conducted on their smartphones. This represents a significant increase over 2019 [15].

When surveyed in this study, patients found the Wellinks platform to be both easy to use and valuable. These 2 factors have been found to be among the highest predictors of intent to use a medical app within a senior population, as found in a recent study of 364 older adults with an average age of 75 years [23]. In our study, the 2 factors with the lowest scores were the

ability to learn about COPD via the app, and enhanced connection to the physician. While access to educational content and 2-way communication are planned for the next-generation Wellinks platform, the study version of the app did not provide any educational material, nor was there any provision for direct communication between the patient and the physician. Patients or their caregivers had the opportunity to give the interviewer feedback about the Wellinks Platform at the end of the study period. Some patients reported technical difficulties, or the need for more clear instructions; others suggested ideas for improvement, such as integration of educational content to teach patients about the device readings and what the numbers meant. Several patients expressed an appreciation for the ability to track spirometry and oximetry values in real time.

NPS has proven to be a remarkably powerful assessment of customer satisfaction across multiple industries [24] and is now increasingly used in the health care setting due to its ease of administration and higher response rates. NPS surveys can be readied more quickly than email or phone surveys and are less time-consuming for patients, and thus more likely to garner a response [25]. NPS methodology is based on research that cross-references survey responses with actual consumer behavior, demonstrating that a strong personal recommendation of a company or a service to one's family and friends was a key measure of satisfaction [26], and directly correlated with uptake in the market [19]. An NPS of 59, as seen with the Wellinks platform, indicates a very high level of patient satisfaction [19].

Limitations

First, it must be noted that mobile and eHealth innovations are often subject to a high rate of attrition, for various reasons; for example, the novelty wears off over time, or patients have difficulty fitting the technology into their everyday lives. This can make it difficult to assess these types of technologies [27,28]. In the case of the Wellinks platform, the base requirement for spirometry and oximetry was only 1 use/week. Therefore, we believe the decline in usage for these 2 parameters is more reflective of hyperutilization toward the beginning of the study which then reverted to the anticipated baseline. The reduction in use of the medications feature may be reflective of the high medication burden for patients with COPD in general. Yet, even the reduced usage rate represents more information than would normally be available to clinicians.

Second, this study was a small pilot with all patients selected by 1 physician at 1 pulmonology practice. Certain study parameters limited our study population, for example, the requirements of access to an iPhone and of patient nebulizer

use. The fact that the study recruited from 1 small geographic area limits its generalizability as it may not reflect a diverse population.

Third, patient engagement may be artificially higher than expected in a real-world situation due to the Hawthorne effect (where individuals modify their behavior because they are being observed) [29]. In this study, patients were recruited by the principal investigator and were aware that their use of apps and monitoring devices was being recorded. This may have motivated the patients to use the app more consistently than if they were not aware that they were being observed. The slow decline in usage over 8 weeks may reflect that effect over time.

Lastly, while useful data regarding patient health status were provided via the patient report, there was no formal requirement for ongoing physician assessment of the data provided. The study was not designed to show improvement in COPD, only that patients would utilize the Wellinks system. Consequently, there were no interventions taken, and the study was not powered to show improvement in clinical outcomes or pharmacoeconomic impact.

Future Directions

This important pilot study proves that the multimodal nature of the Wellinks Platform is feasible and that appropriate utilization and engagement by the target population is possible. The information collected will direct the continued development of the Wellinks Platform, including features such as virtual pulmonary rehabilitation and health coaching, and has the potential to create value for patients and the health care system. Important questions remain: can this platform improve the quality of care for patients with COPD? Can it improve clinical outcomes such as a reduction of hospital readmissions and cost and improving patient quality of life? These questions will be addressed in a planned future clinical study.

Conclusions

The Wellinks platform is a novel platform designed for patients with COPD that consolidates multimodal biometric data and patient entries into a single, streamlined source of data including trends over time and is intended for use by both patients and providers. Patients were eager to engage with the platform and found it to be useful and valuable. Patient engagement was high across disease severity, age, and gender groupings. Given the difficulty of managing this complex, chronic disease, the Wellinks platform has the potential to improve patient engagement and outcomes, ultimately reducing COPD-related events and expenses.

Authors' Contributions

Both authors contributed to the study and preparation of this manuscript.

XSL•FO

RenderX

Conflicts of Interest

BDG is a scientific advisor to Wellinks. CRR is an operating partner of HighCape Capital with a financial interest in Wellinks.

Multimedia Appendix 1

Responses to the Patient Satisfaction questionnaire (n=16). Patients submitted an answer on a 5-point scale ranging from strongly disagree to strongly agree.

[DOCX File , 14 KB - formative_v6i3e34758_app1.docx]

References

1. CDC basics about COPD - Chronic obstructive pulmonary disease (COPD). Centers for Disease Control and Prevention. 2021 Jun 09. URL: https://www.cdc.gov/copd/basics-about.html [accessed 2021-09-15]
2. COPD Prevalence. American Lung Association. 2021. URL: https://www.lung.org/lung-health-diseases/lung-disease-lookup/copd/learn-about-copd [accessed 2021-09-15]
3. Xu JQ, Murphy SL, Kochanek KD, Arias E. Mortality in the United States, 2018. NCHS data brief, no 355. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nchs/data/databriefs/db355-h.pdf [accessed 2021-09-15]
4. Ford ES, Murphy LB, Khavjou O, Giles WH, Holt JB, Croft JB. Total and state-specific medical and absenteeism costs of COPD among adults aged ≥ 18 years in the United States for 2010 and projections through 2020. Chest 2015 Jan;147(1):31-45. [doi: 10.1378/chest.14-0972] [Medline: 25058738]
5. COPD Costs. Centers for Disease Control and Prevention. 2018 Feb 21. URL: https://www.cdc.gov/copd/infographics/copd-costs.html [accessed 2021-10-07]
6. Shah T, Churpek MM, Coca Perraillon M, Konetzka RT. Understanding why patients with COPD get readmitted: a large national study to delineate the Medicare population for the readmissions penalty expansion. Chest 2015 May;147(5):1219-1226 [FREE Full text] [doi: 10.1378/chest.14-2181] [Medline: 25539483]
7. Criner GJ, Bourbeau J, Diekemper RL, Ouellette DR, Goodridge D, Hernandez P, et al. Prevention of acute exacerbations of COPD: American College of Chest Physicians and Canadian Thoracic Society Guideline. Chest 2015 Apr;147(4):894-942 [FREE Full text] [doi: 10.1378/chest.14-1676] [Medline: 25321320]
8. Shah SA, Velardo C, Farmer A, Tarassenko L. Exacerbations in Chronic Obstructive Pulmonary Disease: Identification and Prediction Using a Digital Health System. J Med Internet Res 2017 Mar 07;19(3):e69 [FREE Full text] [doi: 10.2196/jmir.7207] [Medline: 28270380]
9. Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease 2021 Report. GOLD. 2021. URL: https://goldcopd.org/ [accessed 2021-09-10]
10. McCabe C, McCann M, Brady AM. Computer and mobile technology interventions for self-management in chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2017 May 23;5:CD011425 [FREE Full text] [doi: 10.1002/14651858.CD011425.pub2] [Medline: 28535331]
11. Fan KG, Mandel J, Agnihotri P, Tai-Seale M. Remote Patient Monitoring Technologies for Predicting Chronic Obstructive Pulmonary Disease Exacerbations: Review and Comparison. JMIR Mhealth Uhealth 2020 May 21;8(5):e16147 [FREE Full text] [doi: 10.2196/16147] [Medline: 32348262]
12. Abraham WT, Bensimhon D, Pinney SP, Feitell SC, Peacock WF, Amir O, et al. Patient monitoring across the spectrum of heart failure disease management 10 years after the CHAMPION trial. ESC Heart Fail 2021 Oct;8(5):3472-3482 [FREE Full text] [doi: 10.1002/ehf2.13550] [Medline: 34390219]
13. Salehi S, Olyaeemanesh A, Mobinizadeh M, Nasli-Esfahani E, Riazi H. Assessment of remote patient monitoring (RPM) systems for patients with type 2 diabetes: a systematic review and meta-analysis. J Diabetes Metab Disord 2020 Jun;19(1):115-127 [FREE Full text] [doi: 10.1007/s40200-019-00482-3] [Medline: 32550161]
14. Gelburd R. Examining the state of telehealth during the COVID-19 pandemic. United Hospital Fund. 2020 Jun 29. URL: https://uhfnyc.org/publications/publication/telehealth-during-covid-19 [accessed 2021-10-28]
15. 2020 Tech Trends and the 50+: Top Ten Biggest Trends [Slides]. AARP Research. 2021 Apr. URL: https://www.aarp.org/content/dam/aarp/research/surveys_statistics/technology/2021/2021-tech-trends-older-adults.doi.10.26419-2Fres.00420.001.pdf [accessed 2021-10-08]
16. Desai NR, Diamond EJ. Emerging Role of Remote Patient Monitoring in Pulmonary Care: Telemedicine to Smart Phone. Chest 2021 Feb;159(2):477-478. [doi: 10.1016/j.chest.2020.10.015] [Medline: 33563433]
17. Cooper CB, Sirichana W, Arnold MT, Neufeld EV, Taylor M, Wang X, et al. Remote Patient Monitoring for the Detection of COPD Exacerbations. Int J Chron Obstruct Pulmon Dis 2020;15:2005-2013 [FREE Full text] [doi: 10.2147/COPD.S256907] [Medline: 33061338]
18. NPS calculator: calculate your net promoter score. Delighted, LLC. 2021. URL: https://delighted.com/nps-calculator [accessed 2021-10-06]
19. Reichheld FF. The one number you need to grow. Harv Bus Rev 2003 Dec;81(12):46-54, 124. [Medline: 14712543]
20. Hegewald MJ, Gallo HM, Wilson EL. Accuracy and Quality of Spirometry in Primary Care Offices. Annals ATS 2016 Dec;13(12):2119-2124. [doi: 10.1513/annalsats.201605-418oc]

XSL•FO

RenderX

21. Liistro G, Vanwelde C, Vincken W, Vandevoorde J, Verleden G, Buffels J, COPD Advisory Board. Technical and functional assessment of 10 office spirometers: A multicenter comparative study. Chest 2006 Sep;130(3):657-665. [doi: 10.1378/chest.130.3.657] [Medline: 16963659]

22. Roberts ET, Mehrotra A. Assessment of Disparities in Digital Access Among Medicare Beneficiaries and Implications for Telemedicine. JAMA Intern Med 2020 Oct 01;180(10):1386-1389 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2666] [Medline: 32744601]

23. Askari M, Klaver NS, van Gestel TJ, van de Klundert J. Intention to use Medical Apps Among Older Adults in the Netherlands: Cross-Sectional Study. J Med Internet Res 2020 Sep 04;22(9):e18080 [FREE Full text] [doi: 10.2196/18080] [Medline: 32624465]

24. Companies the use net promoter. Net Promoter System. 2021. URL: https://www.netpromotersystem.com/about/companies-that-use-net-promoter [accessed 2021-10-26]

25. Alismail A, Schaeffer B, Oh A, Hamiduzzaman S, Daher N, Song H, et al. The Use of the Net Promoter Score (NPS) in an outpatient allergy and pulmonary clinic: an innovative look into using tablet-based tool vs traditional survey method. PROM 2020 May;11:137-142. [doi: 10.2147/prom.s248431]

26. Hamilton DF, Lane JV, Gaston P, Patton JT, Macdonald DJ, Simpson AHRW, et al. Assessing treatment outcomes using a single question: the net promoter score. Bone Joint J 2014 May;96-B(5):622-628. [doi: 10.1302/0301-620X.96B5.32434] [Medline: 24788496]

27. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

28. Beukenhorst AL, Howells K, Cook L, McBeth J, O'Neill TW, Parkes MJ, et al. Engagement and Participant Experiences With Consumer Smartwatches for Health Research: Longitudinal, Observational Feasibility Study. JMIR Mhealth Uhealth 2020 Jan 29;8(1):e14368 [FREE Full text] [doi: 10.2196/14368] [Medline: 32012078]

29. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol 2014 Mar;67(3):267-277 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.08.015] [Medline: 24275499]

Abbreviations

AARP: American Association for Retired People

COPD: chronic obstructive pulmonary disease

FEV$_1$: forced expiratory volume in 1 second

FEV$_1$%: forced expiratory volume in 1 second as % of predicted for the patient

GOLD: Global Initiative for Chronic Obstructive Lung Disease

HIPAA: Health Insurance Portability and Accountability Act

IRB: Institutional Review Board

NPS: Net Promoter Score

PRN: pro re nata (as needed)

SpO$_2$: peripheral oxygen saturation

XSL•FO

RenderX

Original Paper

Paving the Way for Electronic Patient-Centered Measurement in Team-Based Primary Care: Integrated Knowledge Translation Approach

Selena Davis[1], PhD; Marcy Antonio[2], PhD; Mindy Smith[3,4], MD; Paul Burgener[5], PEng; Danielle C Lavallee[6], PhD; Morgan Price[7], MD, PhD; Sarah C Fletcher[7], PhD; Francis Lau[1], PhD

[1]School of Health Information Science, University of Victoria, Victoria, BC, Canada

[2]School of Information, University of Michigan, Ann Arbor, MI, United States

[3]College of Human Medicine, Michigan State University, Grand Rapids, MI, United States

[4]Patient Advisory Committee, Kootenay Boundary Collaborative Services Committee, Castlegar, BC, Canada

[5]Patient Voices Network, BC Patient Safety & Quality Council, Vancouver, BC, Canada

[6]Michael Smith Health Research BC, Vancouver, BC, Canada

[7]Innovation and Support Unit, Department of Family Practice, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Selena Davis, PhD
School of Health Information Science
University of Victoria
PO Box 1700 STN CSC
Victoria, BC, V8W2Y2
Canada
Phone: 1 2505057866
Email: daviss@uvic.ca

Abstract

Background: Patient-centered measurement (PCM) aims to improve the overall quality of care through the collection and sharing of patient values, outcomes, and perspectives. However, the use of PCM in care team decisions remains limited. Integrated knowledge translation (IKT) offers a collaborative, adaptive approach to explore best practices for incorporating PCM into primary care practices by involving knowledge users, including patients and providers, in the exploratory process.

Objective: This study aims to test the feasibility of using patient-generated data in team-based care; describe the use of these data for team-based mental health care; and summarize patient and provider care experiences with PCM.

Methods: We conducted a multi-method exploratory study in a rural team-based primary care clinic using IKT to co-design, implement, and evaluate the use of PCM in team-based mental health care. Care pathways, workflows, and quality improvement activities were adjusted iteratively to improve integration efforts. Patient and provider experiences were evaluated using individual interviews relating to the use of PCM and patient portals in practice. All meeting notes, interview summaries, and emails were analyzed to create a narrative evaluation.

Results: During co-design, a care workflow was developed to incorporate electronically collected patient-generated data from the patient portal into the electronic medical record, and customized educational tools and resources were added. During implementation, care pathways and patient workflows for PCM were developed. Patients found portal use easy, educational, and validating, but data entries were not used during care visits. Providers saw the portal as extra work, and the lack of portal and electronic medical record integration was a major barrier. The IKT approach was invaluable for addressing workflow changes and understanding the ongoing barriers to PCM use and quality improvement.

Conclusions: Although the culture toward using PCM is changing, the use of PCM during care has not been successful. Patients felt validated and supported through portal use and could be empowered to bring these data to their visits. Training, modeling, and adaptable PCM methods are required before PCM can be integrated into routine care.

(JMIR Form Res 2022;6(3):e33584) doi:10.2196/33584

Introduction

Background

Seminal work in the 1960s and 1970s, supporting the combination of the concepts of a medical home with attributes of patient-centered care [1], provided the foundation for research that demonstrated the benefits of this care model (eg, comprehensive, coordinated care within a primary care team). Benefits linked to patient-centered medical home models include improved health-related quality of life, self-management, and depression; reduced hospital admissions; and improved clinical measures such as blood pressure and glycated hemoglobin [2]. However, the concept of patient-centeredness has been poorly theorized and operationalized, although several papers have identified key attributes such as patient and family being respected; given complete health information; being involved in decision-making; and supported in their physical, psychological, and social needs [3,4]. Newly realized in British Columbia (BC), the patient's medical home is a community practice that operates at an ideal level to provide longitudinal patient-centered, team-based primary care [5].

To assess these attributes, measures are needed to capture patient-centered care in a form that can be used to rate care quality and quality improvement (QI). Patient-centered measurement (PCM) has the potential to capture data to improve patient care experiences, care quality, communication, and trust. To this end, the BC Ministry of Health and the 7 health authorities established the BC PCM Steering Committee [6] in 2021 to implement scientifically rigorous approaches to collect and report patient-generated data (PGD). Examples of sources of these data include patient-reported experience measures (PREMs) and patient-reported outcome measures (PROMs).

However, the integration of PCM into regular care visits and decisions presents major conceptual, methodological, and logistical challenges in translating this body of knowledge into routine clinical practice [7-9]. A systematic review found that PGD improved patient health awareness and communication with providers but that difficulties arose when patients wanted greater provider involvement with their data during clinic visits [10]. Lordon et al [10] found that providers had difficulty accommodating patient requests for engagement with PGD because of the perceived lack of value, time constraints, and lack of workflow integration. Even with access to patient portals for data entry, the ability to incorporate and track these data varies across systems because of organizational, practice, workflow, resource, and technological challenges [11-13].

There is an unprecedented opportunity to develop and test best practices for incorporating PCM into clinical care because of the rapid uptake of virtual care and enhanced digital literacy for both patients and providers during the COVID-19 pandemic. This requires planning to determine which strategies to use, selection considerations about which patient populations to target for PREMs and PROMs, choice of specific measures, and engagement considerations, including how clinical care teams will incorporate and act on the data [14]. Given the complexity of implementing PCM in clinical practice, multilevel implementation science frameworks are effective, with the choice of framework or theory based on *fit for purpose* [15,16]. Iterative knowledge sharing, or integrated knowledge translation (IKT), was identified for our research, whereby planning, implementing, and evaluating team-based care performance could be optimally developed through the lens of providers, patients, and the research team [17]. Synergies derived from this IKT approach enhance the understanding of patients' and providers' context and needs, thereby enhancing the relevance of the generated research and increasing user knowledge and understanding of the research process, awareness of the research, and appreciation for how and when it can be applied [18].

For this research, we partnered with the regional district of Kootenay Boundary (KB) Division of Family Practice [19], whose mission is to help rural practitioners meet patient and practice needs and lead change as part of a province-wide initiative to strengthen health care in BC. The focus on people living in rural and remote communities is critical to address limited access to providers and services, and the resulting health disparities of higher chronic disease multimorbidity and all-cause mortality [20-22]. In this context, digital health solutions have great potential to address these health disparities [23]. The study team assessed priority areas for rural clinical care and identified mental health care as a high-needs area based on BC Community Health Data [24]. Mental illness is a common and disabling health problem in Canada, affecting 1 in 5 Canadians, and has become of even greater importance during the COVID-19 pandemic [25]. Care team–based patient-centered planning strategies have important potential in the treatment of mental illnesses, such as anxiety and depression [26]. Mental health concerns have worsened during the pandemic; based on the 2019 Community Health Survey, almost 5 million Canadians aged ≥12 years (16%) had seen or spoken to a health care professional in the previous year about their mental health, an increase of 2% since 2015 [27].

Objectives

The overall research aim was to develop new methods to incorporate patient-generated mental health and experience data for team-based in-clinic and virtual care. The new methods for PCM that emerged from this study are reported in another publication (M Antonio et al, unpublished data, December 2021). This paper reports on the IKT approach that was used in the multi-method study with the following objectives: (1) to explore the feasibility of integrating PGD using a patient portal in team-based care; (2) to describe the use of these data for team-based mental health care; and (3) to summarize patient and provider care experiences with PCM.

Methods

Setting

The study took place in a rural team-based care practice in the southern interior of BC, Canada, between February 2020 and April 2021. The private, multi-provider primary care clinic served as the patient's medical home [5]. The study received institutional ethical approval (protocol number: BC H19-03855). Notably, the first stage began at the onset of the COVID-19 pandemic, and the study was completed during the pandemic. As such, the research was conducted virtually, including regular interactions with providers, patients, and the research team.

Study Design

The research used a multi-method, IKT approach with interconnected stages of study (Figure 1). In line with the

directive in a scoping review on IKT in health care [28], we created a protocol, timeline, and IKT plan to guide our efforts, and the analysis of continuous and documented feedback offered the research team the ability to report findings with sufficient detail to reveal how IKT was associated with outcomes. Stage 1 of the study involved co-design during which the research team worked with the care team to identify relevant PCMs and optimize the use of the patient portal in the context of the team's clinical needs, roles and responsibilities, and workflow. Stage 2 included the implementation of the portal and care workflows with adaptation based on feedback and evaluation of its use and impact on clinical care. Implementation and evaluation occurred concurrently so that feedback could be incorporated and further evaluated.

Figure 1. Multi-method, integrated knowledge translation approach for the integration of patient-generated data.
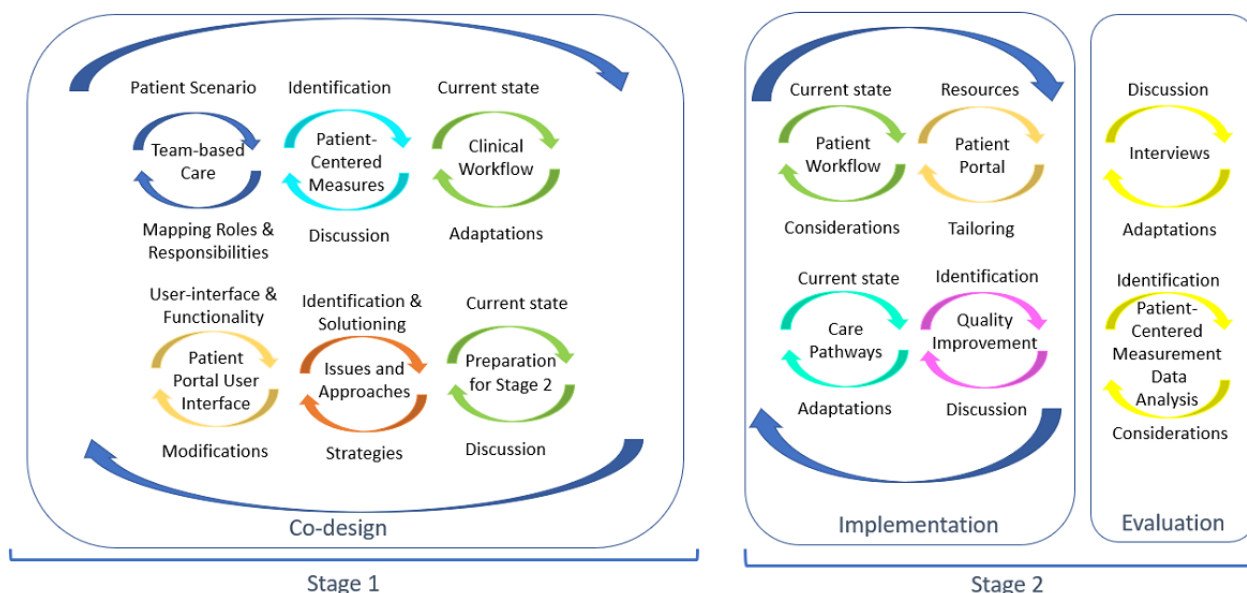
The research team included patient partners, researchers, subject-matter expert scholars, and our industry partner. The patient partners (PB and MS) and two researchers (MA and SD) met regularly with clinic staff and care providers recruited to the study and with other knowledge users at various points within the study, including local professionals from our study collaborators, KB Division of Family Practice and General Practice Service Committee Practice Support Program [29]. Monthly research meetings were held with the entire research team to discuss and apply what we had learned to date.

Recruitment and Composition of the Provider Practice

A pragmatic approach and convenience sampling were used to recruit a community team-based clinic. The KB Division of Family Practice invited one of the early adopter clinics of a patient medical home care model to participate in both study stages. The enrolled clinic included 2 physicians, 2 medical office assistants, and a newly hired social worker and registered nurse (RN). The practice used an electronic medical record (EMR) but had no experience with the use of a patient portal, and their EMR did not have an embedded portal. The practice

had an identified patient panel within the EMR and received summary patient data from the KB Patient Experience Survey (a questionnaire currently in use by the Divisions of Family Practice to understand patient experience with care) and the Canadian Primary Care Sentinel Surveillance Network [30], a primary care research network that offers a web-based data presentation tool to improve primary health care delivery outcomes across the country. However, these data were not used to practice QI.

One of the initial joint decisions was the identification of the clinical domain of study for this intervention—mental health care. The factors considered in this decision included the limited number of mental health providers in rural primary care, the high prevalence of mental illness [25], and the providers' current familiarity with and routine use (including at the point of care) of two patient-generated mental health measures—the Patient Health Questionnaire-9 (PHQ-9) depression measure [31] and the Generalized Anxiety Disorder 7 item (GAD-7) [32].

Co-design Stage 1

Overview

This stage involved coproducing the implementation and evaluation plans. The components are shown in Figure 1. Methods for data collection included team mapping, care workflow, portal development, emails, and summary meeting notes. Outputs of this stage included mapping the care team's roles and responsibilities, identifying relevant PCMs, detailing the office and clinical care workflows to enable PCM use, optimizing the use of the patient portal, and documenting learnings to understand issues and generate solutions.

Team-Based Care Mapping

Team roles and responsibilities were mapped using the team mapping method, a facilitated cocreation workshop designed to help groups explore how to work together in a primary care team [33] and informed by the circle of care modeling [34], an analytical technique to develop the PCM adoption methods and create and validate patient use cases. A 2-hour–long team mapping session, using patient personas focused on mental health concerns, was held over Zoom (Zoom Video Communications) with all providers and staff from the clinical practice and members of the research team. The mapping session explored and defined team roles and responsibilities in caring for the simulated patients and how PCMs could support care decisions. This exercise also explored the use of technology, such as patient portals outside of regular visits to engage patients, and workflow associated with monitoring the data. Following the session, facilitators generated a summary report that included images of the patient-centered maps that were created, as well as summaries of the PCM gaps and potential solutions, categorized for discussion with the care team.

Patient-Centered Measurements

Discussions with providers led to the following selected PREM and PROMs: (1) a multipart question from the KB Patient Experience Survey that contained subitems about whether someone from the care team talked with patients about difficulties taking care of their health, main health goals and priorities, stressors, needed support, medication review, and offer of preventive care; (2) Patient-Reported Outcomes Measurement Information System Item Bank (version 1.0)—General Self-Efficacy Short Form 4a—a general population measure with demonstrated good convergent validity, internal consistency reliability, model fit, and sufficient unidimensionality [35,36]; (3) PHQ-9 depression measure; (4) GAD-7; and (5) self-action plan for depression [37]. The PHQ-9 and GAD-7 are validated measures already in use by the practice at the point of care.

Care Workflow

A group session, held with the clinic after the team mapping exercise, focused on reviewing the mapping report and how PCM gaps could be addressed through alternative workflows [38]. An electronic patient-reported outcomes toolkit [39] guided discussions related to improvements in office efficiency and integration of electronic PGD. Iterations of care workflow were brought forward as a flow diagram for feedback at the next meeting. In consultation with patient partners and co-designed

with the care providers and clinic staff, an initial workflow was identified for the implementation stage.

Patient Portal

The research team selected a commercial vendor portal that had a patient-centered perspective (eg, patients could select who they want to share data within their care team). The web-based portal was used as a stand-alone system during the study (ie, was not interoperable with the clinic's EMR system) as no current tethered or interoperable patient portal was available for use in this study, and the industry partner was amenable to adding selected PCMs and tailoring to study needs (eg, education and resources). Ongoing adjustments to the user interface were gathered and provided to the industry partner for implementation.

The portal had a user interface for patients and a separate one for the clinic staff and care providers. The researchers had access to both the interfaces. Of the 7 functionalities assessed in a recent Cochrane review on patient access to EMRs [40], this portal provided for PCM, including tracking, education, and a reminder feature that was sent if a requested patient questionnaire was not completed within 3 months. The portal did not provide the ability to request other information, bidirectional communication and sharing, or the ability of patients to manage their care. The materials provided in the portal included selected PCMs and educational resources.

Preparing for Implementation and Evaluation

Preparing for stage 2 encompassed patient recruitment, change management support, portal deployment, and development of the evaluation.

Recruitment of Patient Participants

Clinic staff and care providers contacted patients who had been living with a mental health diagnosis for many years to inform them about the study. Interested patients were provided with a flyer that included contact information for a researcher (MA) who provided study information and a consent form before enrollment. Four patients agreed to participate in the study. Participation involved using the patient portal over 4 months and interviews to discuss how PCM and educational materials within the portal were used during their care. Owing to the COVID-19 pandemic, most clinical visits between patients and providers were held over phone.

Implementation—Stage 2

This stage comprised activities to operationalize the intervention and adapt design components based on iterative feedback. The components of this stage are shown in Figure 1.

Patient Workflow and Portal Use

The use of the portal was presented to patients at the start of the intervention as a part of their workflow. Iterative patient workflow adaptations were made based on patient feedback during the implementation and evaluation stages.

One researcher (MA) provided training in portal use for all patients, clinic staff, and care providers. The patient user interface allowed patients to see messages from their provider, complete the PREM and PROMs, and explore educational

materials. The portal was set up to send an email to inform patients that there was an invite in their portal account. Both patient and provider portal interfaces allowed completed PCMs to be viewed on the computer device screen or downloaded as a document that clinic staff could upload into the patient's record in their EMR. Patients and providers could also view data trends of these measures (eg, scores on PHQ-9) if more than one was completed.

Care Workflow for Providers

The care workflow, finalized in the previous stage, was used as the initial care workflow for the study implementation. Individual discussions with providers and clinic staff during the intervention focused on changes to the care workflow specific to care pathways such as screening, monitoring, and follow-up resulting from the addition of PCM to team-based care processes. For example, the RN took responsibility for sending PROMs and educational materials to patients biweekly to complete through the portal. The PREM was sent at the beginning and end of the portal-use period.

We recognized that because of the lack of interoperability between the portal and EMR, we would have to simulate some embedded portal functions, such as sending reminders and data entry. For example, the portal-completed PROMs were intended to be reviewed during clinic visits. To achieve this, the RN and a researcher (MA) shared responsibility for checking for upcoming appointments, deploying the PROM and checking for completion and transferring the PROM score into the EMR encounter notes section, attaching the document, and entering the score into the existing PROM sections of the EMR. The researchers met periodically with the RN to adjust the implementation efforts.

Evaluation Plan—Stage 2

Interviews

A general interview guide approach was used for the patients (Multimedia Appendix 1) and providers (Multimedia Appendix 2). Patient interview questions included interactions with the portal, use of questionnaire results and educational materials by themselves and with the provider, and how other tools could be used to complement communication with the portal. Interviews were scheduled within 1-5 days after a virtual or phone visit with their provider. A total of 13 patient interviews with 4 patients were conducted, and the number of interviews with each patient was dependent on the number of care visits (2-4 visits) during the study period. Provider questions included

visit information, use of patient questionnaires during the visit, whether educational materials prompted discussion, and whether the questionnaire results/educational materials influenced the visit or team interactions. An individual interview was conducted with all 4 care providers and 2 clinic staff members at a convenient time following a patient visit (timing varied between 1 week and 1 month). To gain immediate insights, a web-based survey was provided to physicians immediately after a patient visit (Multimedia Appendix 3).

All interviews were conducted virtually (phone or video) by two researchers (SD and MA) and lasted 10 to 30 minutes. Both researchers had experience with qualitative research. Notes were taken by both interviewers and combined as summary notes. The summary interview notes were coded by one researcher (MA) using ATLAS.ti (ATLAS.ti Scientific Software Development GmbH), and the coding reports and original data were reviewed by 2 additional team members with experience in the analysis of qualitative data (MS and SD). The final interview coding reports were confirmed by the research team.

PCM Data Analysis

The framework method for the analysis of qualitative PCM data was used. The framework method applies a matrix structure to facilitate the recognition of patterns and has been used effectively under the leadership of experienced qualitative researchers [41]. Data analyses involved looking across interview coding reports, team mapping reports, portal use, and care workflow diagrams to determine areas of convergence or divergence and develop a narrative of the evaluation of the implementation. Our reflective practice comprised the development of new care workflow diagrams, iterative writing, and discussion among the research team at monthly analysis meetings.

Results

Co-design—Stage 1

The collective results of stage 1 were used to inform stage 2.

Team Mapping to Inform Care Workflow

The team mapping exercise focused on two areas: care of patients with mental health concerns and how PCM could be used in practice. The session produced *circle of care* maps with defined team roles and responsibilities in caring for the simulated patients (Figure 2). The session's discussion focused on PCM integration and how PCM could influence care.

Figure 2. Mapping team-based care for patients with mental health issues.
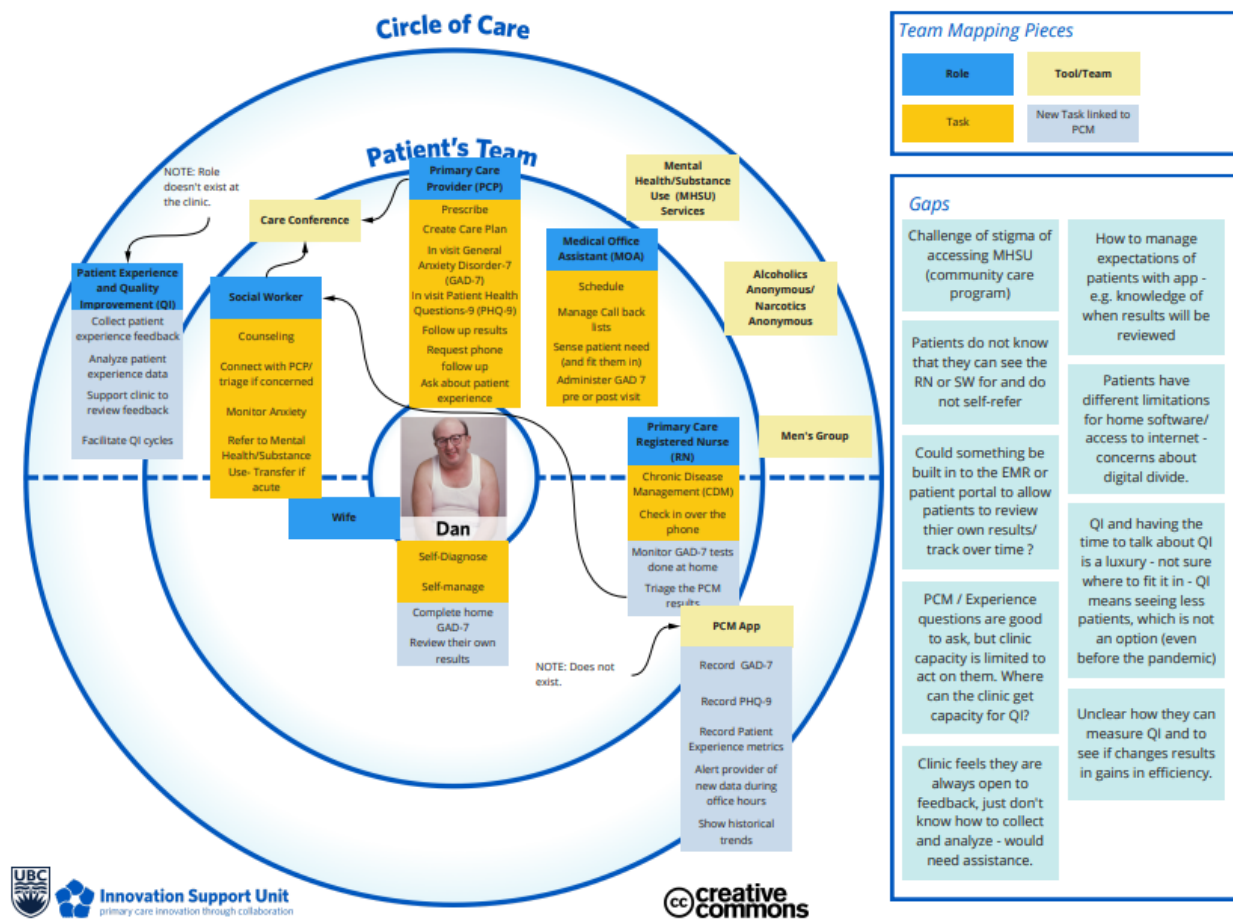
Supplementary File 5: Team Map- Completed team map of persona "Dan", an accountant with history of anxiety and alcohol mis-use. Note the different roles in Dan's circle of care from himself, family, primary care team, as well as other external groups. Team mapping allowed participants to collaboratively discuss features for integrating patient-centred measurement.

Following the mapping session, co-designing efforts produced an initial care workflow for incorporating PREM and PROMs into routine care, before, during, and in between a patient's clinic visits, using a patient portal. Five key activities were identified: (1) deploying PCM, (2) collecting electronic patient data, (3) tracking completion, (4) reviewing data, and (5) documentation. Workflow changes were then designed to incorporate these activities, and a final workflow diagram was developed (Figure 3 [42]). Before the visit, the patient or clinic staff may initiate a visit appointment, the care team may tailor resources and PCM questionnaires in the portal, the patient would receive notifications and review the resources, and then complete the questionnaires. The nurse would track and triage the PCM scores and initiate urgent responses if needed. During the virtual or in-person visit, the physician might review the PCM scores with the patient, document appropriate actions, and refer the patient to other providers as needed. In between visits, the care team would comanage the patient.

Figure 3. Final team-based care workflow with electronic patient-generated data [42]. GAD-7: Generalized Anxiety Disorder 7-item; PHQ-9: Patient Health Questionnaire-9; PROM: patient-reported outcome measure; PROMIS: Patient-Reported Outcomes Measurement Information System.
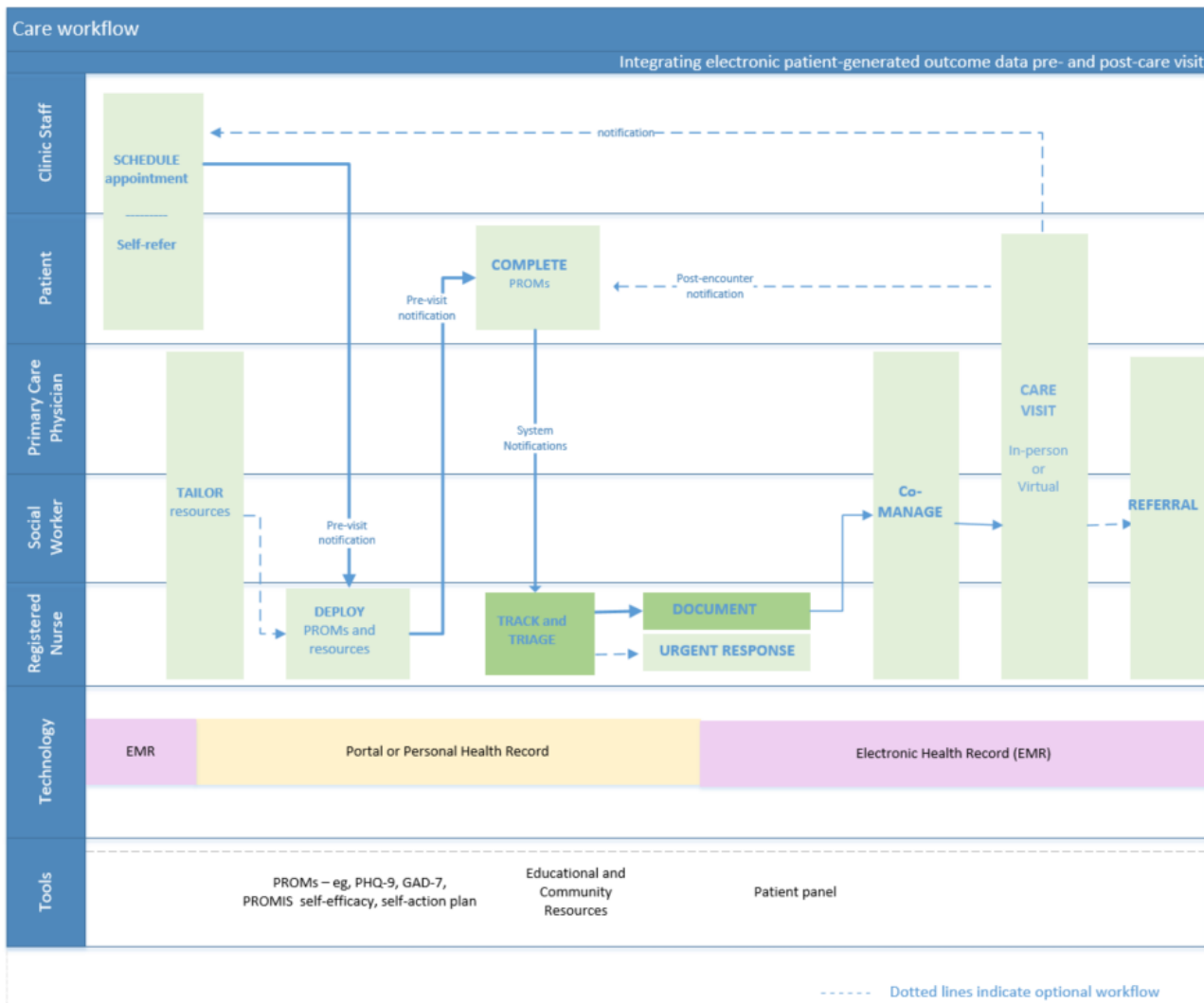
Strategies and Concerns About Integrating PCM

On the basis of team mapping and care workflow sessions, we constructed Table 1 with learnings and potential solutions to care gaps and team limitations, strategies needed to allow the use of PCM, constraints in the current workflow including restrictions because of the COVID-19 pandemic, and modifications needed to the patient portal to enhance care. Providers commented that the mapping and workflow sessions mirrored clinical practice well. The major concerns expressed included the practice's limited capacity for incorporating PCM gathered outside the clinic visit (both electronic and paper-based), patient expectations for review of their data, and the provider's knowledge and clinic capacity that limited engagement in QI. Providers expressed concerns about the rural context with lack of access to high-speed internet, insufficient resources to manage practice workload, and current fee-for-service structures that did not provide an allowance for the collection and use of PCM.

Table 1. Stage 1—gaps, learnings, and potential solutions before implementation.

Co-design step	Learnings	Potential solutions
Team roles and responsibilities	• Patients do not know about roles or access to a RN[a] or social worker • No individual was trained in using data for QI[b] • The stigma around mental health care and access to community program • No ability to track medications or change in PGD[c] scores	• Information about staff and other resources in portal • Engage with training resources (eg, future staff training) • Improve connections so these are part of team • Expand team: pharmacist
Strategies (what needs to be in place to use PGD)	• Patients and providers need access to data and tracking; best if integrated • Patients may need reminders: bring data or trigger discussion • Increase the level of patient engagement	• Make PGD available through the portal; EMR[d] linkage; explore ways to document • Consider reminders in the portal • Provide resources in the portal
Care workflow	• No capacity or training to use patient experience data for QI • COVID-19 pandemic restrictions prevent hallway conversations • There is no way to trigger follow-up (eg, significant change in PGD score)	• Begin with joint selection of an experience measure • Schedule brief virtual work huddles at the start of each day • Plan to explore options in stage 2
Patient portal	There was concern regarding limited digital literacy • There was concern regarding patient expectations for review of data • Patient knowledge gaps around depression and anxiety	• Patient interview data to understand needs • Careful use of language and messages to patients on expected follow-up • Add educational resources to portal and delivery plan

[a]RN: registered nurse.

[b]QI: quality index.

[c]PGD: patient-generated data.

[d]EMR: electronic medical record.

Customization of the Patient Portal

The patient portal was collaboratively prepared with selected mental health questionnaires, patient experience measures, depression self-action planning tool, and educational resources, and comprised a feature for trending data such as the anxiety questionnaire, GAD-7, in Figure 4. An appraisal guide was created to select the educational and resource material to ensure that the material was evidence-based, used patient-oriented language, and was adaptable to local contexts (eg, local community resources). Examples of the educational materials are provided in Multimedia Appendix 4.

Figure 4. The patient portal—patient-centered measurement questionnaires, resources, and functions.

Before deployment, the portal underwent initial testing by patient partners and an undergraduate student who simulated the role of the patient. Through discussions with the practice and research team, the type, timing of deployment, quantity of educational material, messaging, and notifications within the portal were determined before launching portal use.

Implementation and Evaluation—Stage 2

Care Pathways and Portal Use

On the basis of workflow discussions with care providers and patient partners, a clinical care pathways diagram was produced (Figure 5 [42]) to identify the opportunities for care team members to integrate PCM (eg, asynchronous screening or follow-up or during an encounter). There are 7 PCM-related clinical care activities performed by different care team members to screen, triage, assess, diagnose, treat, monitor, and follow up with the patient, which can be guided by their PCM data. Nurses and social workers may initiate screening, monitoring, and triage of patients, or when triggered by web-based notifications and reminders from the portal. Physicians may refer to PCM data when assessing, diagnosing, and treating the patient during an in-person or virtual visit. In between visits, the nurse could follow up with the patient to provide resources and PCM questionnaires within the portal, monitor for completed PCM data, and schedule notifications as reminders to complete them.

Figure 5. Clinical care pathways with electronic patient-generated data [42]. EMR: electronic medical record; GAD-7: Generalized Anxiety Disorder 7-item; PHQ-9: Patient Health Questionnaire-9; PROM: patient-reported outcome measure.
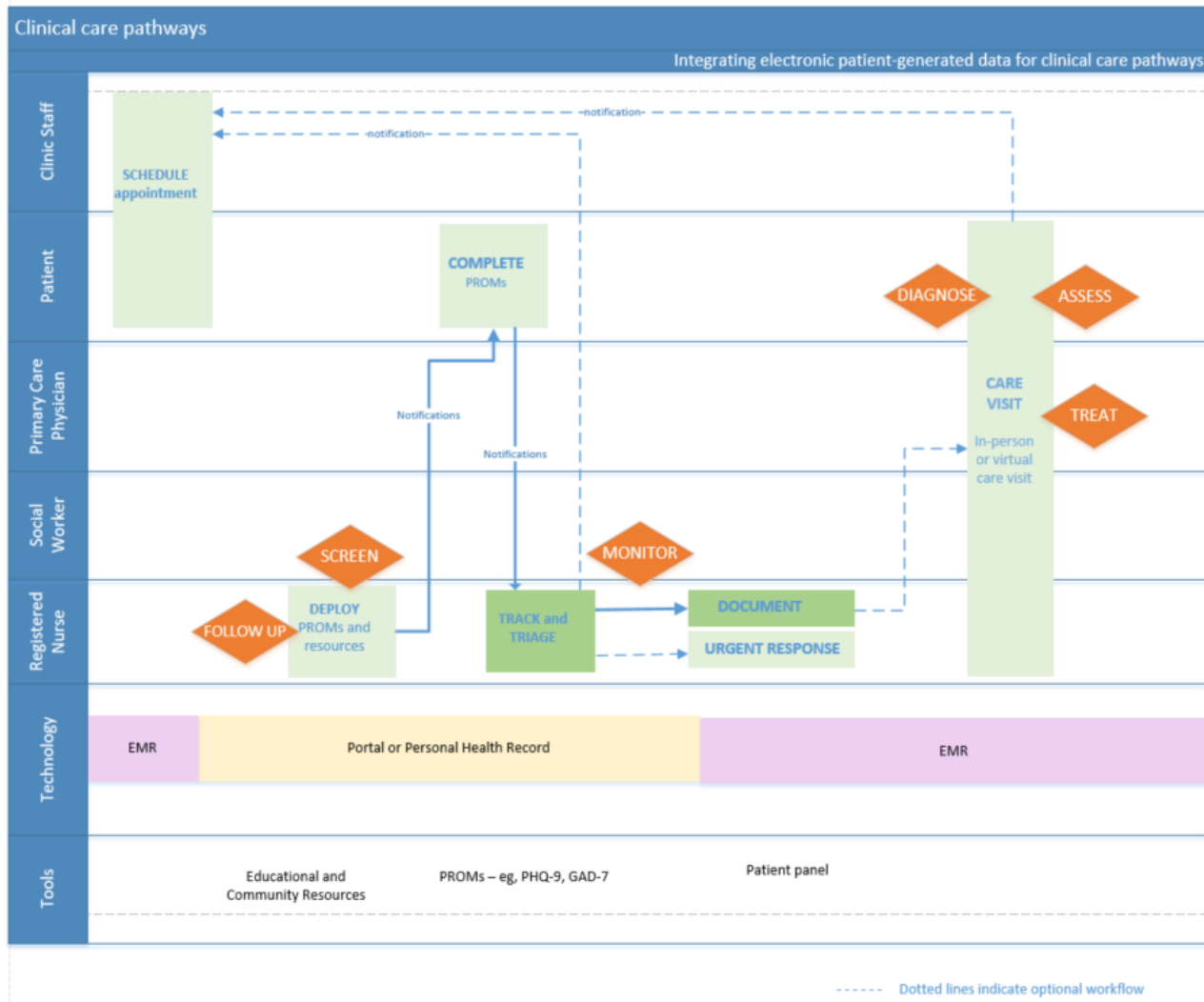
A patient workflow based on patient feedback was created to incorporate real and potential actions related to PCM (Figure 6 [42]). PCM-related activities for patients are to review resources in the portal and evaluate their relevance, respond to notifications, complete PCM questionnaires indicated by the provider or on their own initiative, and take actions in response to new learnings. These actions include reflecting on their condition in response to new information, initiating self-care, tracking trends over time, sharing information with family or care network, reviewing additional resources, following up with the care team, and preparing for a visit. Considerations included delivery-method notification, the timing for completion of questionnaires and resources, tailoring efforts of educational materials, and visit type (in-person or virtual). A video was created [43] to illustrate study results relating to the use and integration of PGD into clinical care from the patient's perspective.

Figure 6. Patient workflow with electronic patient-generated data [42]. PHR: personal health record; PREM: patient-reported experience measure; PROM: patient-reported outcome measure.
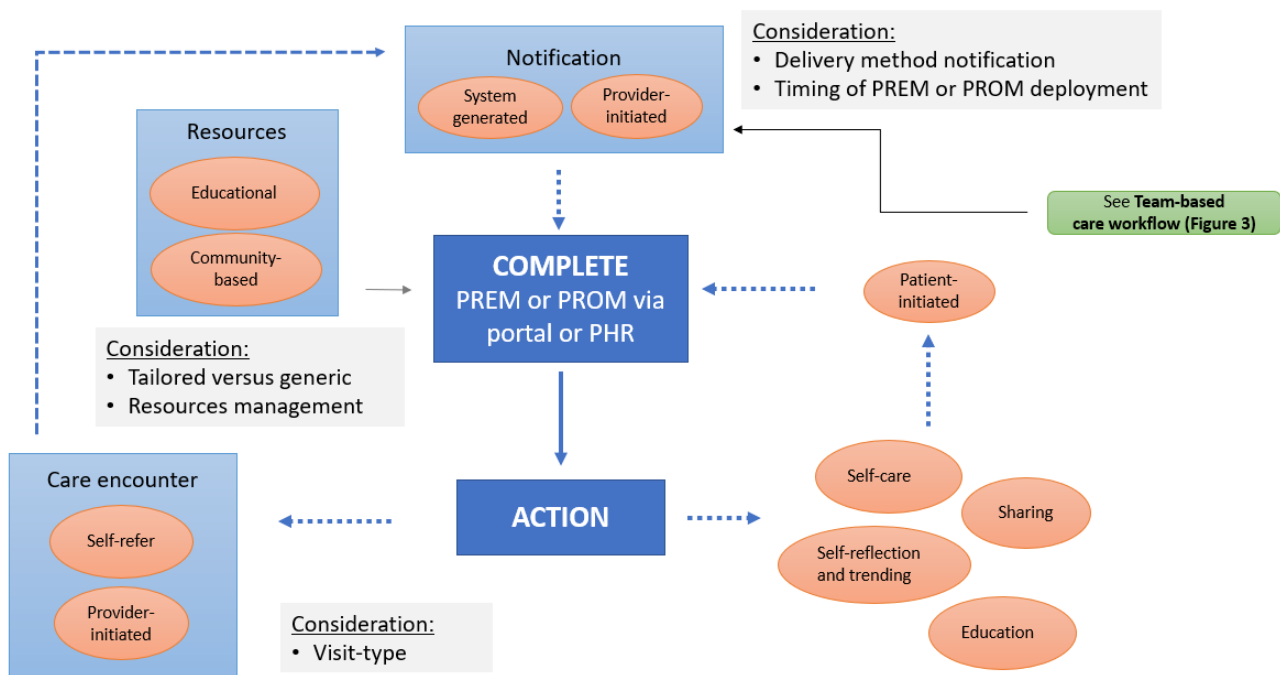

Table 2 provides excerpts from the summary notes that demonstrate our key learnings from stage 2. The qualitative synthesis of 13 patient interviews indicated that the portal was easy to use and valued. Two patients noted the inability to use the portal during times of severe depression and fatigue. Providers saw benefits in patients' use of the portal for information and to give them a sense that care was nearby. The major problems identified with provider use of the portal were lack of portal integration with the EMR and lack of alerts for changes in questionnaire scores that should trigger action. For providers, to limit the extra work across the team, 1 team member (RN) oversaw the PCM data in the portal and transferred it into the EMR. Even with the simulated integration of data transfer into the EMR by the RN, the process of reviewing those data elements in the EMR was often overlooked during a care visit.

All patients read at least some of the educational material provided, with skill-building workbooks being the most appreciated. One patient stated that "knowledge of our illness is vital to taking steps to live with illness and attain victory as often as possible." Getting information over time rather than all at once was useful to avoid overwhelming the participant. Most patient participants had been living with a mental health diagnosis for many years and reported that the initial educational materials offered would have been helpful earlier in their diagnosis and illness. At their request for more advanced information, patient-friendly abstracts of scientific papers were added to the portal. Patients also wanted the ability to rate articles for their usefulness. Providers appreciated the potential for educational resources to be tailored to the patient's stage and needs. However, they recognized that they would not have the resources or domain expertise needed to sustain ongoing educational material development or deployment in the portal.

Table 2. Stage 2: learnings from patient-centered measurement (PCM) and portal implementation.

Learnings	Patients	Providers
Use of patient portal (not used during visits)	• View educational and community resources and complete PROMs[a] and PREMs[b] • A way to overcome isolation and focus on self • Unable to use during times of severe depression or fatigue	• Ways to provide information to the patient and impart "a feeling that care is nearby" • A lack of integration with EMR[c] • Resulted in use by a single nurse to manually transfer patient-generated data, which were rarely reviewed during a care visit • Lack of alerts to trigger action reduced portal usefulness
Value of educational resources	• Knowledge is vital to overcoming illness • Provide credible information and avoid getting "lost in the abyss" • The initial educational materials provided would have been helpful earlier in their diagnosis and illness. Personalized material would be valuable	• Appreciated the potential for educational resources to be tailored to the patient's stage and needs • "Not able to sustain ongoing educational material development or deployment"
Use of PCMs	• Completing PROMs was extremely "validating" • Sense of being heard and capturing more of the relevant information about their mental health • Tracking and trending scores using a portal "painted a picture of where I am" • Frustration that their providers were not asking about or reviewing PROMs during care visits • Unclear who was to bring up the PROM	Did not use within portal because of the following reasons: • Lack of integration of portal with EMR (only total score of PROMs were manually entered into the EMR) • Lack of alerts for changes in questionnaire scores that should trigger action • Belief that the PCM did not add to the existing relationship
Optimizing completion and use of PROMs	• Timely reminders for PROM completion • PROMs needed that address function to aid "what created my responses" as part of the interpretation of a measure	• Timely customizable reminders for PROM completion • Interoperability of systems • Additional training on use of PROMs

[a]PROM: patient-reported outcome measure.

[b]PREM: patient-reported experience measure.

[c]EMR: electronic medical record.

PCM Use in Team-Based Mental Health Care

Three patients reported that the process of completing PROMs was extremely *validating* in that they learned more about themselves and gained confidence in reaching out for support. As one patient noted on reviewing the PHQ-9 with a resident physician, the experience was "unexpectedly rewarding and validating (same page together) and resulted in a treatment change." Several patients noted that clinic visits in their current form are often used to revisit past appointments or begin with general questions that may not relate to a patient's current status. As 1 patient, who was hesitant to talk with the physician about anxiety, noted, completing the questionnaire ensured that all pertinent aspects were covered, preventing communication disconnect. In addition, tracking scores were appreciated; 1 patient stated that the weekly trend "painted a picture of where I am." However, 1 patient noted difficulty in interpreting depression scores in that they did not provide insight into "what created my responses." Suggested additions by patients for PROMs included functional measures (eg, what is working now for or against you), more sophisticated graphics for tracking scores and more specific questions in the action plan based on depression scores. Patients wanted to be able to complete PROMs on their own and trend data over multiple years.

Patients expressed frustration that their providers did not review PROMs and did not ask about PCM at visits. One patient

reported that she thought she was doing something wrong in not having questionnaire data at her visit. Providers noted that although PCM collected before a visit could make visits more efficient, they did not review them for reasons noted in Table 2.

As PGD in the portal was a new activity, there was also a lack of clarity when the PROM scores would be viewed and discussed. Patients were used for in-person visits and thus did not consider opening the portal during their virtual visit. Each provider had different preferences for where to view these data but had not established a practice of viewing them. The limited-time for visits meant that providers prioritized the purpose of the visit and planned to look at the PROM scores during the patient's annual mental health check-up.

Concerning optimizing the use of PCMs, both patients and providers wanted to be able to set timely reminders for PROM completion. Patients sometimes logged into the portal to see an invite and start completing a PROM, and then realized that they wanted to return when they could give more focus or were less fatigued. The lack of reminders meant that patients often forgot to return and complete PROM. Similarly, providers wanted the ability to customize reminders (eg, 24 hours before a visit or a week after the initial invite). Providers also noted that they lacked education and training on PROMs' limited use, including

interpretation of a PCM score and what constituted a significant change.

Quality Improvement

Although QI activities were never completed, discussions were held to explore opportunities within workflow and care processes to carry it out. Examples of mental health care included PHQ-9 completion, counseling, and medication prescriptions for patients with a diagnosis of depression. The resultant QI process workflow is displayed in Figure 7 [42]. PCM-related QI activities for the care team are based on the plan-do-study-act cycle and documentation processes that can

be informed by PCM. Planning is driven by clinic priorities and capacities and is further guided by community health profiles and the latest clinical practice guidelines. Doing involves running patient panel queries using the clinic EMR and related external databases or applications to identify areas for attention. Studying, acting, and documentation involves a detailed review of the panel query outputs, implementing specific actions, and documenting the actions and results when available. The PCM focus brings awareness to specific groups of patients that may require action depending on their health situation and PCM scores.

Figure 7. Quality improvement (QI) in team-based care workflow with electronic patient-generated data [42]. EMR: electronic medical record; PREM: patient-reported experience measure; PROM: patient-reported outcome measure.
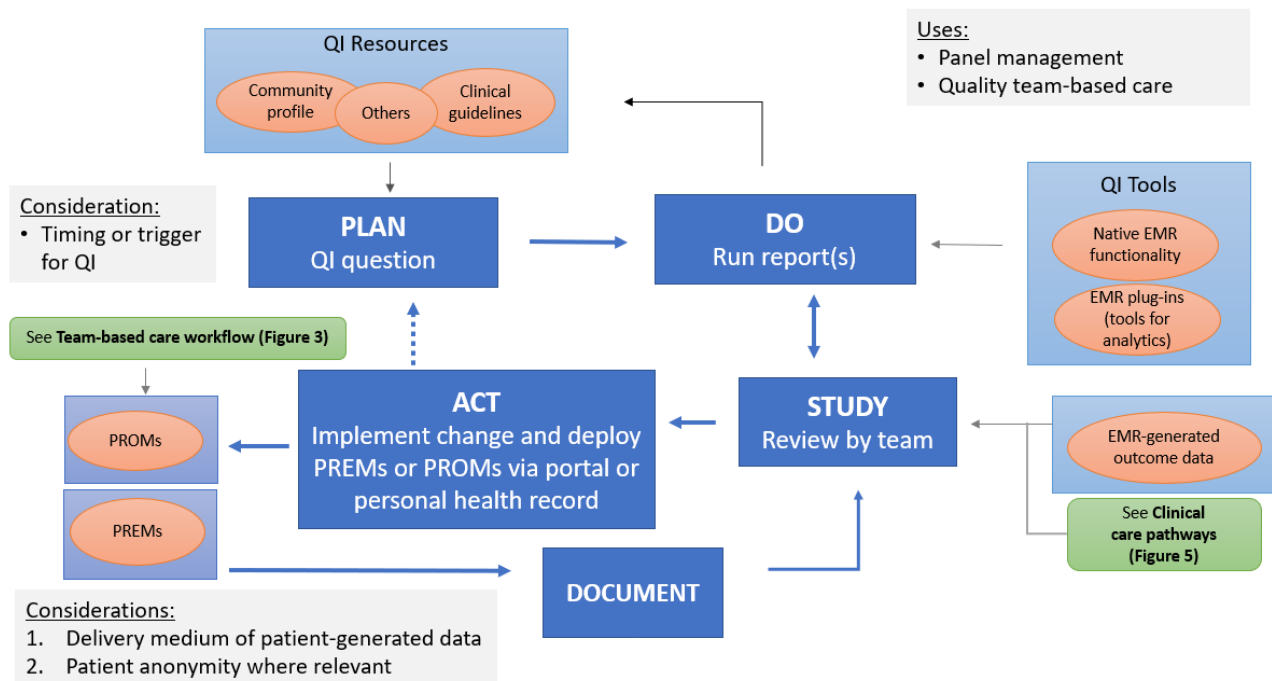
To date, most QI efforts have been focused at the regional level, and the clinic does not know how to apply PREM results to clinical processes with multiple providers. An agreement was received around its value, but resistance was observed, and one provider commented on the lack of medical school training in QI. There was also a sense of technology overload, where providers had to remember multiple login codes and how to view data across more than 4 different systems. QI tools and resources available, delivery medium and type of PCM, timing and triggers for QI, and maintenance of patient anonymity when using PREMs were discussed.

Additional Learnings

Knowledge sharing (or IKT) throughout the study was invaluable to our process and learnings. Specifically, during all interactions and interviews with knowledge users, the researchers received productive feedback to questions like *how we are doing* and whether *we hit the mark* with iterations of (1) process descriptions; (2) design planning, implementation, and evaluation actions; and (3) documentation of identified issues and solutions.

Although providers noted that it takes a *whole team* to care for patients with mental health disorders, clinical resources were overwhelmed and there were limited community mental health services in this rural area. Multiple patients commented on how their physician was overworked, so they were mindful of what could be done at a visit, although they would have preferred mental health visits more frequently than once annually. The addition to the practice of the social worker was relatively recent, and one patient did not know about the availability of the RN or social worker or how they could fit into their care. Another patient stated that they wanted a physiotherapist to be part of their team. Both patients and providers noted that rural living was associated with fewer specialist resources and longer wait times.

Since the COVID-19 pandemic, virtual care visits have become a routine part of practice. One staff member noted that patients were no longer in the waiting room or available to complete questionnaires on-site. Both patients and providers felt that virtual care was less valuable because of lack of personal contact; one provider stated that "the loss of human connection is devastating." One patient reported that they might have brought up their anxiety if face-to-face and that being virtual

missed the seriousness of their illness. However, patients also preferred to answer PROMs at home, as the answers seemed more honest than completion in a public waiting room. One clinic staff member noted a possible benefit of phone visits in a rural (small population) clinic in that patients could have a degree of anonymity by not having a face to recognize when seen out in the community.

Discussion

Principal Findings

IKT offered the research a continuous and rich understanding of patient and provider needs and current challenges in the context of team-based care and PCM, as well as potential best practices for integrating PGD using a patient portal. Providers were unable to incorporate the use of electronically generated mental health PROMs within team-based care during this study for several reasons including workload, need to prioritize issues to address during an encounter, lack of easy access to the data, lack of value placed on the data, and lack of education about and practice in its use. The absence of established provincial standards for interoperability between systems hampered the integration of electronic PCMs in team-based care. In a review of personal health record functionalities and implementation, Harahap et al [44] identified interoperability as a key implementation issue, as well as security and privacy, usability, data quality, and personalization as other important factors. Digital health transformation at the system level is needed to realize interoperability, providing standard definitions for data exchange and cooperation with the patient, provider, and organizational systems. Using an ontological information model, Plastiras and O'Sullivan [45] demonstrated the feasibility of transferring PGD using common standards from personal health records to EMRs. Interoperability must be addressed to ensure that electronic PROMs and PREMs from patient systems can be fully integrated within provider systems and readily available at the point of care. In addition to the issue of interoperability, the authors perceived that the limited experience of these newly formed primary care teams, and having to shift to virtual care during the pandemic made it less likely that any member could ensure delegation and use of the data provided.

Solutions to learnings from the co-design step that were acted upon (ie, engaging with the local Division of Family Practice around QI, patient reminders for the completion of questionnaires, research assistant tracking, and providing patient education and resources within the portal) were valued and demonstrative of the way forward. Ultimately, the gap in getting data viewed in the EMR and usable at the time of a visit was not accomplished. For digital tools to be successful in addressing provider workload, authors of a systematic review [40] identified a need for training, reducing documentation and task time, expanding the care team, and leveraging QI processes in workflows.

Although QI was an intended part of PCM advancement, it was discovered early in the study that providers felt that the practice could not engage in practice improvement using PREMs. An additional knowledge user with subject-matter expertise in QI was added to the team to present information on EMR panel report analysis, community profile, best practices, and clinical practice QI guidelines using PROMs. Although providers expressed interest in the future use of PCM for QI, because of the COVID-19 pandemic, frustrations expressed over prior poor quality QI data provided, and inexperience with use, the study shifted to provide an opportunity to explore QI and educate the practice on ways PCM could be used for QI. The QI activities were not implemented or evaluated.

PREMs are being collected through population surveys in the geographic area of this study, but QI efforts are still nascent, and there is no single best way to collect or use PCM for QI. Gleeson et al [46] systematically reviewed current QI efforts using PREMs and found that most practices attempted small, incremental changes to services that did not require a change in provider behavior and resulted in unclear impacts. They called for more attention to how PREM data can be used to inform practice changes that have a positive influence on the patient care experience. Translating new ideas into practice among early-career providers requires three considerations—credibility, practicality, and need [47]. Efforts to incorporate QI into routine practice will likely require attention to messaging and creating digital solutions to address these issues. In addition, simply providing PGD, even if entered into the EMR, has the potential to increase provider burnout. A review of health record-integrated PCM found that technostress (technological complexity, uncertainty, overload, insecurity, and invasion [higher patient expectations]), time pressure, and workflow-related issues need to be addressed to accelerate the integration of PCM into clinical care [48]. Therefore, future endeavors will need to consider the human and fiscal resources needed for QI in clinical practice and the integration of PGD into the digital health ecosystem.

To create primary care team-based practices that value and use PCM to improve care, a culture change is needed. This process, in the experience of one of the authors (MS) in teaching evidence-based medicine and research training, takes 3 to 5 years, with time for training and practice through targeted, small-scale projects. For example, the introduction of QI training of multidisciplinary teams in local health departments in North Carolina resulted in small but important changes in organizations' cultures over 5 years, increasing engagement in future QI, and improving overall care and services [49]. Mandating QI may be problematic, with a before-and-after study of Foundation Year 1 doctors in the United Kingdom reporting less motivation to complete QI projects and placing less importance on QI for their professional development despite a significant increase in overall QI knowledge at the end of the year [50]. It remains to be seen if the BC support system approach to practice facilitation for QI is successful in this effort.

However, patients found the process of completing and tracking the results of mental health-related questionnaires on portal validation useful in capturing relevant aspects of their experiences, a unique feature of this study. The importance of portals as a communication tool and limited use of PGD was noted in an umbrella review [51], and lack of bidirectional communication was likely an additional barrier to PCM used in this study. Patients commented on how PROMs could provide a way to focus visits on what is of the greatest value for them

at the moment. PCM may be particularly relevant for patients with illnesses for which there are no biomedical markers. For people living with mental health concerns, these measures can aid in communication when the illness makes it difficult to communicate the severity of their difficulties. Although evidence supports the use of patient medical home care models that include data-driven quality of care and patient engagement [2], what seems to be missing is the value placed on PCM by providers and an understanding and appreciation of ways in which the use of these data can advance communication, care quality and QI. To achieve the patient-centered potential of PCM requires a conceptual shift in workflows, where patients and providers are encouraged to bring PGD, particularly around function, into care discussions.

Study Limitations

As this was an exploratory study within a single clinic, the transferability of the findings to another context needs to be explored further. The study was carried out in a rural, early adopter patient medical home practice with a small care team and few patient participants, which may not be representative of other primary care models, larger clinics, or urban settings. As the study focused on mental health, the relevance of our findings to other health conditions is limited.

The researchers supported transferability by providing a detailed description of the context and location, and trustworthiness by being transparent about our methods for data collection and analysis. The study was conducted during the pandemic, which made recruitment of care teams and patients particularly challenging. We strove to address the limited sample size by adding richness to our data through multiple study methods and

an iterative process of knowledge sharing between researchers, including patient partners and knowledge users. Being a transdisciplinary team of patient partners, providers, and researchers across multiple disciplines enabled us to bring in multiple perspectives during the analysis.

Limitations were introduced during the data collection and analysis. The persona image was selected to reflect how Dan may be dressed in an examination room. Upon reflection, we realize that Dan's appearance would be quite different during a virtual visit and would have selected a persona image that does not reinforce power differentials or stereotyping [52]. Although we did not record interviews, we used 2 interviewers at each session, and summative interview reports were iteratively compared with all PCM study data by the research team. Memos in the summary notes were used to record when the interviewers had different interpretations. Each researcher established their process for reflection, and we did not have a standard way to record and review individual reflections.

Conclusions

The value of PGD and the need for PCM methods to collect, integrate, and use PGD in team-based care remain challenging. Through collaborative and adaptive efforts, this gap was narrowed by demonstrating ways in which PGD can be incorporated into clinical practice within a Canadian team-based primary care setting. The conceptualization of PCM methods to accomplish this goal is well served by an IKT approach. IKT offers a beneficial technique for addressing our knowledge of users' needs related to the collection, integration, and use of PGD, bridging the knowledge to action gap and setting the stage for future success.

Authors' Contributions

FL proposed the research and drafted the initial grant proposal. All authors have contributed to the revised and final grant proposal. SD led the engagement with the provider participants. MA led the engagement with patient participants and industry and academic partners. PB, MS, and SD tested and selected material for the portal, and MA integrated education materials and measures into the portal. SD and MA collected interview data. MA and MS performed the analysis of interview data and theme development with input from SD, FL, and PB. SD created Figure 1 and Figures 3-6. SD and MS created the tables. SD and MS drafted the manuscript. All authors critically revised and approved the submitted manuscript and agreed to be accountable for ensuring its integrity.

Conflicts of Interest

A research team member was an industry partner (Cambian Business Services Inc) for this study, who provided data storage and inclusion of the survey tools and educational resources into the patient portal. The system was provided at no cost to the clinic for the duration of the study.

Multimedia Appendix 1

Interview guide and questions for patients.
[DOCX File , 28 KB - formative_v6i3e33584_app1.docx]

Multimedia Appendix 2
Interview guide and questions for providers.
[DOCX File , 26 KB - formative_v6i3e33584_app2.docx]

Multimedia Appendix 3
Questions for providers—web-based survey.
[DOCX File , 34 KB - formative_v6i3e33584_app3.docx]

Multimedia Appendix 4
Examples of educational resources made available on patient portal.
[DOCX File , 26 KB - formative_v6i3e33584_app4.docx]

References

1. The patient centered medical home. Robert Graham Center. URL: https://www.aafp.org/dam/AAFP/documents/about_us/initiatives/PCMH.pdf [accessed 2020-11-05]
2. John J, Jani H, Peters K, Agho K, Tannous W. The effectiveness of patient-centred medical home-based models of care versus standard primary care in chronic disease management: A systematic review and meta-analysis of randomised and non-randomised controlled trials. Int J Environ Res Public Health 2020;17(18):50. [doi: 10.21203/rs.3.rs-35151/v1]
3. Giusti A, Nkhoma K, Petrus R, Petersen I, Gwyther L, Farrant L, et al. The empirical evidence underpinning the concept and practice of person-centred care for serious illness: a systematic review. BMJ Glob Health 2020 Dec 10;5(12):e003330 [FREE Full text] [doi: 10.1136/bmjgh-2020-003330] [Medline: 33303515]
4. Scholl I, Zill JM, Härter M, Dirmaier J. An integrative model of patient-centeredness - a systematic review and concept analysis. PLoS One 2014 Sep 17;9(9):e107828 [FREE Full text] [doi: 10.1371/journal.pone.0107828] [Medline: 25229640]
5. System change - patient medical homes. General Practice Service Committee. 2020. URL: https://gpscbc.ca/what-we-do/system-change/patient-medical-homes [accessed 2021-05-01]
6. BC Patient-Centred Measurement. URL: https://www.bcpcm.ca/ [accessed 2021-03-07]
7. Hsiao C, Dymek C, Kim B, Russell B. Advancing the use of patient-reported outcomes in practice: understanding challenges, opportunities, and the potential of health information technology. Qual Life Res 2019 Jun 25;28(6):1575-1583. [doi: 10.1007/s11136-019-02112-0] [Medline: 30684149]
8. Reading MJ, Merrill JA. Converging and diverging needs between patients and providers who are collecting and using patient-generated health data: an integrative review. J Am Med Inform Assoc 2018 Jun 01;25(6):759-771 [FREE Full text] [doi: 10.1093/jamia/ocy006] [Medline: 29471330]
9. Cohen DJ, Keller SR, Hayes GR, Dorr DA, Ash JS, Sittig DF. Integrating patient-generated health data into clinical care settings or clinical decision-making: lessons learned from project healthdesign. JMIR Hum Factors 2016 Oct 19;3(2):e26 [FREE Full text] [doi: 10.2196/humanfactors.5919] [Medline: 27760726]
10. Lordon RJ, Mikles SP, Kneale L, Evans HL, Munson SA, Backonja U, et al. How patient-generated health data and patient-reported outcomes affect patient-clinician relationships: a systematic review. Health Informatics J 2020 Dec 20;26(4):2689-2706 [FREE Full text] [doi: 10.1177/1460458220928184] [Medline: 32567460]
11. West P, Van Kleek M, Giordano R, Weal M, Shadbolt N. Common barriers to the use of patient-generated data across clinical settings. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, QC, Canada p. 1-13. [doi: 10.1145/3173574.3174058]
12. Chung AE, Basch EM. Incorporating the patient's voice into electronic health records through patient-reported outcomes as the "review of systems". J Am Med Inform Assoc 2015 Jul 21;22(4):914-916 [FREE Full text] [doi: 10.1093/jamia/ocu007] [Medline: 25614143]
13. Lavallee DC, Lee JR, Austin E, Bloch R, Lawrence SO, McCall D, et al. mHealth and patient generated health data: stakeholder perspectives on opportunities and barriers for transforming healthcare. Mhealth 2020;6:8 [FREE Full text] [doi: 10.21037/mhealth.2019.09.17] [Medline: 32190619]
14. Gensheimer SG, Wu AW, Snyder CF, PRO-EHR Users' Guide Steering Group. Oh, the places we'll go: patient-reported outcomes and electronic health records. Patient 2018 Dec;11(6):591-598. [doi: 10.1007/s40271-018-0321-9] [Medline: 29968179]
15. Stover AM, Haverman L, van Oers HA, Greenhalgh J, Potter CM, ISOQOL PROMs/PREMs in Clinical Practice Implementation Science Work Group. Using an implementation science approach to implement and evaluate patient-reported outcome measures (PROM) initiatives in routine care settings. Qual Life Res 2021 Nov 10;30(11):3015-3033 [FREE Full text] [doi: 10.1007/s11136-020-02564-9] [Medline: 32651805]

16. Manalili K, Santana MJ, ISOQOL PROMs/PREMs in Clinical Practice Implementation Science Work Group. Using implementation science to inform the integration of electronic patient-reported experience measures (ePREMs) into healthcare quality improvement: description of a theory-based application in primary care. Qual Life Res 2021 Nov 26;30(11):3073-3084. [doi: 10.1007/s11136-020-02588-1] [Medline: 32715388]

17. Graham ID, Kothari A, McCutcheon C, Integrated Knowledge Translation Research Network Project Leads. Moving knowledge into action for more effective practice, programmes and policy: protocol for a research programme on integrated knowledge translation. Implement Sci 2018 Feb 02;13(1):22 [FREE Full text] [doi: 10.1186/s13012-017-0700-y] [Medline: 29394932]

18. Gagliardi AR, Kothari A, Graham ID. Research agenda for integrated knowledge translation (IKT) in healthcare: what we know and do not yet know. J Epidemiol Community Health 2017 Feb 19;71(2):105-106 [FREE Full text] [doi: 10.1136/jech-2016-207743] [Medline: 27647137]

19. Kootenay Boundary Division of Family Practice. URL: https://divisionsbc.ca/kootenay-boundary [accessed 2019-10-18]

20. Rural health services in BC: a policy framework to provide a system of quality care. Cross Sector Policy Discussion Paper - BC. 2015. URL: https://www.health.gov.bc.ca/library/publications/year/2015/rural-health-policy-paper.pdf [accessed 2021-12-07]

21. Roberts KC, Rao DP, Bennett TL, Loukine L, Jayaraman GC. Prevalence and patterns of chronic disease multimorbidity and associated determinants in Canada. Health Promot Chronic Dis Prev Can 2015 Aug;35(6):87-94 [FREE Full text] [doi: 10.24095/hpcdp.35.6.01] [Medline: 26302227]

22. Bosco CI, Oandasan I. Review of family medicine within rural and remote Canada: education, practice, and policy. College of Family Physicians of Canada, Mississauga, ON. 2016. URL: https://www.cfpc.ca/CFPC/media/Resources/Rural-Practice/ARFM_BackgroundPaper_Eng_WEB_FINAL.pdf [accessed 2021-12-03]

23. Zhang X, Hailu B, Tabor D, Gold R, Sayre MH, Sim I, et al. Role of health information technology in addressing health disparities: patient, clinician, and system perspectives. Med Care 2019 Jun;57 Suppl 6 Suppl 2(6):115-120 [FREE Full text] [doi: 10.1097/MLR.0000000000001092] [Medline: 31095049]

24. BC community health data. Provincial Health Services Authority. URL: http://communityhealth.phsa.ca/ [accessed 2019-12-03]

25. Fast facts about mental illness. Canadian Mental Health Association. URL: https://cmha.ca/brochure/fast-facts-about-mental-illness/ [accessed 2021-12-07]

26. Pinho LG, Correia T, Lopes MJ, Fonseca C, Marques MD, Sampaio F, et al. Patient-centered care for people with depression and anxiety: an integrative review protocol. J Pers Med 2021 May 13;11(5):411 [FREE Full text] [doi: 10.3390/jpm11050411] [Medline: 34068307]

27. Canadian Community Health Survey 2019. Statistics Canada. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/200806/dq200806a-eng.htm [accessed 2021-12-08]

28. Gagliardi AR, Berta W, Kothari A, Boyko J, Urquhart R. Integrated knowledge translation (IKT) in health care: a scoping review. Implement Sci 2016 Mar 17;11(1):38 [FREE Full text] [doi: 10.1186/s13012-016-0399-1] [Medline: 26988000]

29. Practice support program. General Practice Service Committee. URL: https://gpscbc.ca/what-we-do/practice-supports/psp [accessed 2019-09-25]

30. Canadian Primary Care Sentinel Surveillance Network (CPCSSN). URL: https://cpcssn.ca/ [accessed 2019-02-09]

31. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

32. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

33. Price M, Bellwood P, Hill TT, Fletcher S. Team mapping: a novel method to help community primary healthcare practices transition to team-based care. Healthc Q 2020 Jan;22(4):33-39. [doi: 10.12927/hcq.2020.26086] [Medline: 32073389]

34. Price M. Circle of care modelling: an approach to assist in reasoning about healthcare change using a patient-centric system. BMC Health Serv Res 2016 Oct 04;16(1):546 [FREE Full text] [doi: 10.1186/s12913-016-1806-7] [Medline: 27716188]

35. Salsman JM, Schalet BD, Merluzzi TV, Park CL, Hahn EA, Snyder MA, et al. Calibration and initial validation of a general self-efficacy item bank and short form for the NIH PROMIS. Qual Life Res 2019 Sep 28;28(9):2513-2523 [FREE Full text] [doi: 10.1007/s11136-019-02198-6] [Medline: 31140041]

36. PROMIS Gender Self-Efficacy - short form. PROMIS Health Organization. 2017. URL: https://www.healthmeasures.net/explore-measurement-systems/promis [accessed 2020-12-03]

37. Depression self-action plan. Adapted from Ted Amann, RN C. URL: https://www.aafp.org/dam/AAFP/documents/patient_care/nrn/depression-self-care-action-plan.pdf [accessed 2020-12-02]

38. Unertl KM, Weinger MB, Johnson KB, Lorenzi NM. Describing and modeling workflow and information flow in chronic disease care. J Am Med Inform Assoc 2009;16(6):826-836 [FREE Full text] [doi: 10.1197/jamia.M3000] [Medline: 19717802]

39. Lerouge C, Austin E, Lee J, Segal C, Sangameswaran S, Hartzler A. ePROs in clinical care?: guidelines and tools for health systems. CERTAIN. 2020. URL: http://epros.becertain.org/ [accessed 2020-09-07]

40. Ammenwerth E, Neyer S, Hörbst A, Mueller G, Siebert U, Schnell-Inderst P. Adult patient access to electronic health records. Cochrane Database Syst Rev 2021 Feb 26;2(2):CD012707. [doi: 10.1002/14651858.CD012707.pub2] [Medline: 33634854]

41. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]

42. Lau F, Davis S, Antonio M. Team-based methods for patient-centred measurement. BC Patient-Centred Measurement Working Group (BCPCMWG) and the BC SUPPORT Unit. URL: https://onlineacademiccommunity.uvic.ca/patientcentredmeasure/ [accessed 2022-02-01]

43. Davis S, Antonio M, Ritland L. Video: Uses and Integration of PGD Into Clinical Care - The Patient Perspective. 2021. URL: https://youtu.be/dREwtq_fkXA [accessed 2022-02-01]

44. Harahap NC, Handayani PW, Hidayanto AN. Functionalities and issues in the implementation of personal health records: systematic review. J Med Internet Res 2021 Jul 21;23(7):e26236 [FREE Full text] [doi: 10.2196/26236] [Medline: 34287210]

45. Plastiras P, O'Sullivan D. Exchanging personal health data with electronic health records: a standardized information model for patient generated health data and observations of daily living. Int J Med Inform 2018 Dec;120:116-125. [doi: 10.1016/j.ijmedinf.2018.10.006] [Medline: 30409336]

46. Gleeson H, Calderon A, Swami V, Deighton J, Wolpert M, Edbrooke-Childs J. Systematic review of approaches to using patient experience data for quality improvement in healthcare settings. BMJ Open 2016 Aug 16;6(8):e011907 [FREE Full text] [doi: 10.1136/bmjopen-2016-011907] [Medline: 27531733]

47. Yelon S, Sheppard L, Sleight D, Ford J. Intention to transfer: how do autonomous professionals become motivated to use new ideas? Perform Improv Q 2008;17(2):103. [doi: 10.1111/j.1937-8327.2004.tb00309.x]

48. Ye J. The impact of electronic health record-integrated patient-generated health data on clinician burnout. J Am Med Inform Assoc 2021 Apr 23;28(5):1051-1056. [doi: 10.1093/jamia/ocab017] [Medline: 33822095]

49. Schaaf EB, Cornett A, Randolph G. Improving culture, one quality improvement project at a time. J Public Health Manag Pract 2018;24(1):57-62. [doi: 10.1097/PHH.0000000000000554] [Medline: 28383343]

50. Bartlett T, Hewertson E, Vassallo M. Culture change and mandating quality improvement. Clin Teach 2019 Feb 10;16(1):30-35. [doi: 10.1111/tct.12758] [Medline: 29318733]

51. Antonio MG, Petrovskaya O, Lau F. The State of Evidence in Patient Portals: Umbrella Review. J Med Internet Res 2020 Nov 11;22(11):e23851. [doi: 10.2196/23851]

52. Cutting K, Hedenborg E. Can personas speak? Biopolitics in design processes. In: Proceedings of the Companion Publication of the Designing Interactive Systems Conference. 2019 Presented at: DIS '19: Designing Interactive Systems Conference 2019; June 23 - 28, 2019; San Diego CA USA p. 153-157 URL: https://dl.acm.org/doi/abs/10.1145/3301019.3323911 [doi: 10.1145/3301019.3323911]

Abbreviations

BC: British Columbia
EMR: electronic medical record
GAD-7: Generalized Anxiety Disorder 7-item
IKT: integrated knowledge translation
KB: Kootenay Boundary
PCM: patient-centered measurement
PGD: patient-generated data
PHQ-9: Patient Health Questionnaire-9
PREM: patient-reported experience measure
PROM: patient-reported outcome measure
QI: quality improvement
RN: registered nurse

XSL•FO
RenderX

XSL•FO

RenderX

Original Paper

Examining the Influence on Perceptions of Endometriosis via Analysis of Social Media Posts: Cross-sectional Study

Julian Matthias Metzler[1], MD; Dimitrios Rafail Kalaitzopoulos[1], MD; Laurin Burla[1], MD; Gabriel Schaer[1], Prof Dr med; Patrick Imesch[1], MD

Department of Gynecology, University Hospital Zurich, Zurich, Switzerland

Corresponding Author:
Julian Matthias Metzler, MD
Department of Gynecology
University Hospital Zurich
Frauenklinikstrasse 10
Zurich, 8091
Switzerland
Phone: 41 442551111
Email: julian.metzler@usz.ch

Abstract

Background: Social media platforms, such as Facebook and Instagram, are increasingly being used to share health-related information by "influencers," regular users, and institutions alike. While patients may benefit in various ways from these interactions, little is known about the types of endometriosis-related information published on social media. As digital opinion leaders influence the perceptions of their followers, physicians need to be aware about ideas and beliefs that are available online, in order to address possible misconceptions and provide optimal patient care.

Objective: The aim of this study was to identify and analyze frequent endometriosis-related discussion topics on social media in order to offer caregivers insight into commonly discussed subject matter and aspects.

Methods: We performed a systematic search using predefined parameters. Using the term "endometriosis" in Facebook's search function and a social media search engine, a list of Facebook pages was generated. A list of Instagram accounts was generated using the terms "endometriosis" and "endo" in Instagram's search function. Pages and accounts in English with 5000 or more followers or likes were included. Nonpublic, unrelated, or inactive pages and accounts were excluded. For each account, the most recent 10 posts were identified and categorized by two independent examiners using qualitative content analysis. User engagement was calculated using the numbers of interactions (ie, shares, likes, and comments) for each post, stratified by the number of followers.

Results: A total of 39 Facebook pages and 43 Instagram accounts with approximately 1.4 million followers were identified. Hospitals and medical centers made up 15% (6/39) of the Facebook pages and 5% (2/43) of the Instagram accounts. Top accounts had up to 111,600 (Facebook) and 41,400 (Instagram) followers. A total of 820 posts were analyzed. On Facebook, most posts were categorized as "awareness" (101/390, 25.9% of posts), "education and research" (71/390, 18.2%), and "promotion" (64/390, 16.4%). On Instagram, the top categories were "inspiration and support" (120/430, 27.9% of posts), "awareness" (72/430, 16.7%), and "personal story" (72/430, 16.7%). The frequency of most categories differed significantly between platforms. User engagement was higher on Instagram than on Facebook (3.20% vs 0.97% of followers per post). On Instagram, the highest percentage of users engaged with posts categorized as "humor" (mean 4.19%, SD 4.53%), "personal story" (mean 3.02%, SD 4.95%), and "inspiration and support" (mean 2.83%, SD 3.08%). On Facebook, posts in the categories "awareness" (mean 2.05%, SD 15.56%), "humor" (mean 0.91%, SD 1.07%), and "inspiration and support" (mean 0.56%, SD 1.37%) induced the most user engagement. Posts made by hospitals and medical centers generated higher user engagement than posts by regular accounts on Facebook (mean 1.44%, SD 1.11% vs mean 0.88%, SD 2.71% of followers per post) and Instagram (mean 3.33%, SD 1.21% vs mean 3.19%, SD 2.52% of followers per post).

Conclusions: Facebook and Instagram are widely used to share endometriosis-related information among a large number of users. Most posts offer inspiration or support, spread awareness about the disease, or cover personal issues. Followers mostly engage with posts with a humoristic, supportive, and awareness-generating nature. Health care providers should be aware about the topics discussed online, as this may lead to an increased understanding of the needs and demands of digitally proficient patients with endometriosis.

Introduction

Endometriosis, defined as the occurrence of endometrium-like tissue outside the uterus, is a chronic, incurable disease affecting about 10% of women [1]. Main symptoms include dysmenorrhea, dyspareunia, chronic pelvic pain, and subfertility. Although it may lead to severe and sustained restrictions in women's private, social, and professional lives, both men and women have little knowledge about the condition [2], and even general practitioners' expertise about the disease is scarce [3]. This dilemma is further aggravated by the high prevalence of dysmenorrhea in adolescents, leading to a trivialization of pelvic pain. These factors mutually reinforce delayed diagnosis, withheld care, and lack of awareness. As women are dissatisfied with the public and medical aid they receive [4,5], it is understandable that they turn to other forms of support.

The internet has become the primary source for health information for people all over the world [6]. In addition to static websites, social media plays an increasing role as a channel for medical information [5]. Platforms such as Facebook and Instagram are used by the majority of US adults [7], and the number of health-focused social media accounts has been skyrocketing, with daily growth rates of up to 28% in 2020 [8]. Patients with endometriosis may benefit from this development in various ways. The process of informing oneself about health topics on social media platforms is private, yet personally tailored, allowing for sharing of personal anecdotes and interactions with others, if desired. Thereby, social and emotional support may be received [5,9]. Support in the form of disease management tips, experiences, and mental health support are valuable subjects among young adults with chronic disease as well as baby boomers and older adults [10,11].

On social media, content can be created by, and shared among, regular users, commercial companies, or nonprofit organizations alike. While any user can share personal or general health-related posts with their friends or with the public via their personal stream, other forms of information sharing have emerged in order to focus on specific topics. Two frequently observed variants are the creation of topic-specific Facebook pages and Instagram accounts. This allows for the promotion and distribution of topic-specific content and the dedicated recruitment of followers.

Regarding endometriosis, several hundred disease-related Facebook pages have been created so far [6]. Recently, "influencers" have increasingly gained importance on most social media channels, blurring the line between user-generated and advertorial content. It has been shown that these digital opinion leaders are able to change the attitudes of their followers, increase acceptance of the information provided, and even influence the intention to buy corresponding products [12].

Even though there is an abundance of endometriosis-related information online, filtering and assessing the quality of information can be challenging, as people make little use of source credibility [13]. Available information on endometriosis is often of low quality, inaccurate, or skewed toward the diagnosis or it conveys negative connotations, while high-quality information is challenging for a lay audience to comprehend [14,15]. These circumstances can induce feelings of fear and helplessness [16].

Nevertheless, internet health information seeking has become increasingly popular, and a growing number of patients with endometriosis can be expected to have acquired considerable knowledge on their condition before a consultation. While this may be challenging at first, it can improve the patient-physician relationship when the patient is able to discuss the information with the physician [17]. Naturally, this requires health care professionals to have an overview of the topics at hand.

While several studies have performed social media content analysis for different gynecological conditions on Instagram and Facebook [18-21], only few focused on endometriosis [5,22,23], and information about popular topics is scarce. Due to this paucity of data, it is unclear which topics prevail on both Instagram and Facebook and whether user engagement differs between platforms.

In this paper, we analyzed the nature of posts shared on endometriosis-related public Facebook pages and Instagram accounts and compared them between platforms, granting an additional tool to improve counseling, address possible misconceptions, and fulfill the patient's expectations to be more engaged in health-related decision-making.

Methods

Design

We performed a cross-sectional analysis of public, endometriosis-related social media posts on Facebook and Instagram between August 3 and 5, 2020. The timing was selected with sufficient time distance from known events such as Endometriosis Awareness Month in March, which could have acted as a confounder. Facebook and Instagram were chosen as data sources for their high popularity [7]. The search was restricted to results in English and those publicly accessible to all Facebook and Instagram users.

Data Selection

In order to identify relevant Facebook pages, we performed a search using the term "endometriosis" in Facebook's search function on August 3, 2020. We deliberately included only Facebook pages, as Facebook groups are often private and require approval by a moderator to gain access.

Additionally, we performed a search on BuzzSumo, a website offering tools and search utilities for content discovery, content research, and identification of influencers across different social media platforms. In contrast to regular popularity ranking by

likes, this website has the capability of ranking relevant Facebook pages by additional metrics such as "engagement," an integrative score including reactions, comments, shares, and likes [24]. Results were structured by relevance and by likes in order to maximize the results.

Of all search results, the *information section* was screened together with the last 10 public posts to assess the scope and aim of each page. Pages were considered relevant when they covered endometriosis as their main topic. When endometriosis was not the main topic (eg, pages focusing on general health or nutrition), a minimum of 10 endometriosis-related posts within the last 2 years were considered necessary to qualify. Relevant pages in English with 5000 or more likes were included for further analysis. This cutoff was defined in order to limit analysis to content with a considerable reach and influence. Pages (ie, accounts) that did not meet the inclusion criteria (ie, endometriosis not the main topic, <10 endometriosis-related posts within 48 months, or <5000 followers) were excluded.

On Instagram, we searched for relevant accounts using the search terms "endometriosis" and "endo." The search was conducted on August 5, 2020. Follower lists of relevant accounts were also screened in order to maximize search results. Account names and number of followers were extracted into a Microsoft Excel spreadsheet and sorted by number of followers. All relevant accounts in English with 5000 or more followers were included.

Of all eligible Facebook pages and Instagram accounts, the 10 most recent wall posts in Facebook or the 10 most recent posts in Instagram were analyzed further.

Content Analysis, Codebook, and Categories

A systematic qualitative content analysis approach was used, as this is a standard process in interactive media content analysis [25,26]. A codebook was developed in order to systematically analyze the content of all relevant posts and to file them into predefined groups. The codebook contained 10 categories or groups together with a short explanation about each category's applicability. For example, the "inspiration and support" category was defined as "all posts containing tips, support (mental, moral, etc), and inspirational texts, quotes, and memes." The 10 content categories were defined as follows: "education and research," "awareness and outreach," "nutrition, food, and diet," "sport and lifestyle," "inspiration and support," "personal story," "patient requests," "scientific inquiries," "humor," and "promotion of product or service." An 11th category (ie, "other") was available if none of these 10 categories seemed suitable. The categories were selected based on a previously published study on social media usage [19]; they were adjusted in order to fit the formulated research questions and to be more aligned with the investigated social media services, since the cited study was performed on Twitter and not Facebook or Instagram. Therefore, the categories "nutrition, food, and diet," "patient requests," "scientific inquiries," and "sport and lifestyle" were introduced to test their occurrence in the data sample, while the categories "celebrity story," "political," and "news" were omitted.

Of all included Facebook pages and Instagram accounts, the 10 most recent endometriosis-related wall posts on Facebook or the 10 most recent posts on Instagram, as of August 5, 2020, were analyzed according to the codebook. Each post was coded as a single topic by two independent coders. Coders received an introduction to the codebook and a training set of 10 posts. If results differed between coders, the content was examined by a third person, and each case was discussed until agreement could be reached. Facebook pages and Instagram accounts were classified as health care professionals or medical centers if (1) their bio or page information section declared this classification or (2) if the content of their posts made this classification explicit. This categorization was conducted by a single reviewer (Reviewer A).

Calculations and Metrics: User Engagement

To measure the impact of posts and the amount of interactions the content earned relative to reach, different metrics were calculated [27]. Total engagement was defined as the sum of all user interactions with a post. Likes, comments, and shares (Facebook only) were added together with equal weighting, as follows:

Total engagement per post = likes $\times$ 1 + comments $\times$ 1 + shares $\times$ 1

As an example, a post with 100 likes, 20 comments, and 10 shares would result in a total engagement of 130. Next, the percentage of account followers engaging with that post was obtained by calculating the engagement rate by reach (ERR) using the following formula:

ERR = total engagement per post / total followers $\times$ 100

Additionally, we calculated the percentage of followers liking, commenting, or sharing a post.

Statistical Analysis: Interobserver Reliability

Statistical analyses were performed with SPSS Statistics for Windows (version 27; IBM Corp). Data were presented as numbers and percentages. For categorical data, chi-square tests and analyses of variance were used. A *P* value below .05 was considered statistically significant.

Chance-adjusted measurement of interobserver reliability was calculated using the Randolph free-marginal multi-rater κ [28], with κ values less than 0.40 considered as "poor," values from 0.40 to 0.75 considered as "intermediate to good," and values above 0.75 considered as "excellent."

Ethical Considerations

According to Swiss law, this is not a research project under the Swiss Human Research Act (Humanforschungsgesetz) and, therefore, no authorization is required.

Results

Pages, Accounts, and Posts

A total of 39 Facebook pages and 43 Instagram accounts were identified, with approximately 1.4 million followers. Top accounts had up to 111,600 (Facebook) and 41,400 (Instagram)

followers, with a median follower number for Facebook and Instagram of 11,800 (IQR 7450-19,750) and 10,200 (IQR 7420-16,900), respectively. Hospitals and medical centers made up 15% (6/39) of the Facebook pages and 5% (2/43) of the Instagram accounts. A total of 820 posts were included for categorization: 390 (47.6%) Facebook posts and 430 (52.4%) Instagram posts. Interrater agreement was 62.8% (245/390) for Facebook, with a free-marginal κ of 0.59 (95% CI 0.54-0.64), showing substantial agreement. For Instagram posts, the interrater agreement was 57.9% (249/430), with a κ of 0.54 (95% CI 0.49-0.59).

Top Categories by Frequency

All pages (ie, accounts) covered more than one topic (ie, published posts in several categories). There was a significant difference in topics between platforms, as shown by categorization frequency. On Facebook, most posts were categorized as "awareness" (101/390, 25.9% of posts),

"education and research" (71/390, 18.2%), and "promotion of product and service" (64/390, 16.4%), with only a few posts addressing "sport and lifestyle" (6/390, 1.5%). As a benchmark, only 1.8% (7/390) of posts were categorized as "other."

On Instagram, the top categories were "inspiration and support" (120/430, 27.9% of posts), "awareness" (72/430, 16.7%), and "personal story" (72/430, 16.7%). Only 0.2% (1/430) of posts were patients asking specific health questions (ie, "patient requests") or calls for scientific study participation (ie, "scientific inquiries"). A total of 6.7% (29/430) of posts were categorized as "other."

Out of 11 categories, 9 (82%) differed significantly in the frequency of their occurrence on Facebook and Instagram, suggesting that different topics were posted depending on the platform. Posts concerning "nutrition, food, and diet" and "sport and lifestyle" were published on both platforms without a significant difference (Table 1).

Table 1. Frequencies of post categories on Instagram as compared to Facebook.

Category	Facebook					Instagram				P value
	Posts (n=390), n (%)	ERR[a] (likes), %[b]	ERR (comments), %[c]	ERR (shares)%[d]	Total ERR, %	Posts (n=430), n (%)	ERR (likes), %[b]	ERR (comments), %[c]	Total ERR, %	
Awareness	101 (25.9)	0.93	0.28	0.84	2.05	72 (16.7)	2.60	0.13	2.73	.001
Education and research	71 (18.2)	0.20	0.07	0.12	0.39	31 (7.2)	2.20	0.14	2.34	<.001
Promotion	64 (16.4)	0.27	0.03	0.03	0.33	49 (11.4)	0.68	0.06	0.74	.04
Inspiration and support	54 (13.8)	0.38	0.05	0.13	0.56	120 (27.9)	2.72	0.11	2.83	.001
Personal story	46 (11.8)	0.26	0.05	0.09	0.40	72 (16.7)	2.79	0.23	3.02	.02
Nutrition, food, and diet	14 (3.6)	0.08	0.15	0.02	0.25	12 (2.8)	1.53	0.15	1.68	.51
Patient requests	11 (2.8)	0.12	0.12	0.01	0.25	1 (0.2)	1.16	0.95	2.11	.002
Scientific inquiries	8 (2.1)	0.08	0.02	0.01	0.11	1 (0.2)	2.00	0.17	2.17	.01
Humor	8 (2.1)	0.54	0.11	0.26	0.91	36 (8.4)	4.04	0.15	4.19	<.001
Other	7 (1.8)	0.16	0.04	0.02	0.22	29 (6.7)	2.32	0.23	2.55	<.001
Sport and lifestyle	6 (1.5)	0.33	0.06	0.05	0.44	7 (1.6)	0.38	0.08	0.46	.92
All categories[e]	390 (100)	0.30	0.09	0.14	0.54	430 (100)	2.04	0.22	2.26	N/A[f]

[a]ERR: engagement rate by reach; ERR = total engagement per post / total followers × 100.

[b]The percentage of followers liking a post from a given category.

[c]The percentage of followers commenting on a post from a given category.

[d]The percentage of followers sharing a post from a given category.

[e]ERR values in this row are mean values of the 11 rows above.

[f]N/A: not applicable; P values in this column compare the category frequency between Facebook and Instagram; since the final row includes frequencies of 100% for both platforms, a P value was not calculated.

When combining both platforms, the top three categories were "inspiration and support" (174/820, 21.2% of all analyzed posts), "awareness" (173/820, 21.1%), and "personal story" (118/820, 14.4%). Infrequent categories included "sport and lifestyle"

(13/820, 1.6%), "patient requests" (12/820, 1.5%), and "scientific requests" (9/820, 1.1%). Table 2 shows the combined frequency of all posts on Instagram and Facebook.

Table 2. Combined frequency of categories on Instagram and Facebook.

Category	Posts (N=820), n (%)
Inspiration and support	174 (21.2)
Awareness	173 (21.1)
Personal story	118 (14.4)
Promotion	113 (13.8)
Education and research	102 (12.4)
Humor	44 (5.4)
Other	36 (4.4)
Nutrition, food, and diet	26 (3.2)
Sport and lifestyle	13 (1.6)
Patient requests	12 (1.5)
Scientific inquiries	9 (1.1)

User Engagement

User engagement was assessed based on 430 Instagram posts from 43 accounts (563,500 total followers) and 390 Facebook posts from 39 pages (831,900 total followers).

On both platforms, posts were typically liked, shared, or commented on by less than 5% of the followers, with a higher percentage of users engaging on Instagram than on Facebook. Across all categories, followers were over 3 times as likely to share or like a post on Instagram as compared to Facebook. On average, 3.20% (SD 3.65%) of followers engaged with a post on Instagram (liking a post: mean 3.01%, SD 3.41%; commenting a post: mean 0.19%, SD 0.37%) as compared to 0.97% (SD 8.02%) on Facebook (liking a post: mean 0.53%, SD 3.18%; commenting a post: mean 0.12%, SD 1.23%; sharing a post: mean 0.31%, SD 3.65%). Posts made by hospitals and medical centers generated higher user engagement than posts by regular accounts on Facebook (mean 1.44%, SD 1.11% vs mean 0.88%, SD 2.71% of followers per post), and Instagram (mean 3.33%, SD 1.21% vs mean 3.19%, SD 2.52% of followers per post).

Top Categories by Engagement

Facebook

On Facebook, the top three categories by engagement were "awareness," "humor," and "inspiration and support." Posts in these categories induced 2.05% (SD 15.56%), 0.91% (SD 1.07%), and 0.56% (SD 1.37%) of users to engage with that post, in contrast to only 0.11% (SD 0.13%) of followers engaging with calls to participate in scientific studies (ie, "scientific inquiries"). Of note, certain categories, such as "humor" or "sport and lifestyle," accounted for a low number of posts, yet generated a lot of engagement (Table 1).

Next, we compared the number of Facebook posts in each category with the engagement rate of each category. A weak correlation was found between category frequency and ERR (R^2=39.2%), showing that posts that induced high engagement tended to be posted more often or vice versa. Posts addressing disease awareness not only made up for 25.9% (101/390) of posts but also generated over 3.8 times more engagement than the average posts, performing at 382% of the expected engagement rate. Table 3 shows the engagement-inducing performance of posts for all categories on Facebook.

Table 3. Performance of Facebook posts by category.

Category	Proportion of mean performance (engagement), %[a]
Overperforming posts	
Awareness	382
Humor	169
Inspiration and support	104
Underperforming posts	
Sport and lifestyle	81
Personal story	75
Education and research	73
Promotion	61
Nutrition, food, and diet	46
Patient requests	46
Other	43
Scientific inquiries	20

[a]100% represents the expected mean engagement across all categories.

Instagram

On Instagram, the top categories were "humor," "personal story," and "inspiration and support," inducing mean user engagement of 4.19% (SD 4.53%), 3.02% (SD 4.95%), and 2.83% (SD 3.08%) of followers, respectively. Interestingly, "sport and lifestyle" scored last, with a mean user engagement of 0.46% (SD 0.64%) of followers (Table 1). Of note, the mean user engagement across all categories (0.54%, SD 0.52% on Facebook; 2.26%, SD 0.99% on Instagram) differed from the ERR described in the User Engagement section, as the latter was weighted for post frequency. As noted before, categories generating higher engagement tended to be posted more frequently or vice versa, leading to a higher ERR in pages and accounts that pursued this strategy. For Instagram, this correlation was weaker than for Facebook (R^2=14% vs 39%). For example, even though humorous posts induced the most engagement on Instagram, they only accounted for 8.4% (36/430) of all posts. Table 4 shows the engagement-inducing performance of posts for all categories on Instagram.

Table 4. Performance of Instagram posts by category.

Category	Proportion of mean performance (engagement), %[a]
Overperforming posts	
Humor	185
Personal story	134
Inspiration and support	125
Awareness	121
Other	113
Education and research	104
Underperforming posts	
Scientific inquiries	96
Patient requests	94
Nutrition, food, and diet	75
Promotion	33
Sport and lifestyle	20

[a]100% represents the expected mean engagement across all categories.

Discussion

Principal Findings

In this paper, we were able to quantify Facebook's and Instagram's popularity with regard to endometriosis. Over 80 pages and accounts with approximately 1.4 million followers were included. By analyzing 820 posts, we were able to specify the diversity of endometriosis-related content that was shared and discussed among Instagram and Facebook users.

Facebook's top three categories—"awareness," "education and research," and "promotion"—may be explained by the intended use of the Facebook pages. Similar to a personal profile, a page enables organizations and other entities to create a public presence, which is visible to everyone on the internet [29]; this makes it a viable tool to generate topic-specific awareness, broadcast educational content, and promote products or services.

Our results are mostly in line with what was recently reported by Towne et al [5]. Even though a direct comparison is limited due to partly different categories, we classified a similar percentage of Facebook posts as "research and education" (18.2%) in our study compared to "education" in Towne et al (21%), and we found a low percentage of posts in the categories "nutrition, food, and diet" (3.6%) in our study compared to "recipe" in Towne et al (0.83%), "humor" in our study (2.1%) compared to "humor" in Towne et al (4.70%), "scientific inquiries" in our study (2.1%) compared to "survey" in Towne et al (0.34%), and "other" in our study (1.8%) compared to "other" in Towne et al (1.10%).

Similarly, Wilson et al [23] found a high percentage of posts on their endometriosis-related Facebook page to have informational content.

As described, certain categories were found to be less common than others; for example, only a few posts contained disease-specific questions from users posed to their followers ("patient requests": Facebook 2.8% of posts; Instagram 0.2% of posts). A possible explanation is that such intimate questions tended to be asked in closed groups and forums or via direct messages rather than publicly. Furthermore, some categories may be underrepresented because of the topics being discussed on different pages: an account's main topic may differ depending on the number of followers. For every account with over 5000 followers, there were numerous accounts with 50 to 500 followers, highlighting that information is spread not only from the top down, but also between smaller groups of peers. Without being fully investigated in this study, these smaller Facebook pages often tended to focus on the personal history or story of a single patient. This may explain a relative underrepresentation of the category "personal story" in the study results.

This is the first study showing that content creators share different topics depending on the platform, as popularity and frequency of topics vary between Facebook and Instagram. While most posts on Facebook were creating educational or promotional awareness, on Instagram, the topics were found to be somewhat more "intimate," with a focus on inspiration and support, awareness, and personal stories. This interpretation may be valid for different reproductive topics as well. When

analyzing fertility-related accounts on Instagram, Blakemore et al [19] found a high percentage of posts containing inspiration and support (24% vs 27.9% in our work) as well as personal stories (32% vs 17.0%), while the topic of education and research appeared less frequently on Instagram posts (11% vs 7.2%). Of note, content may vary due to time or certain events; Gochi et al [22] found significant changes in topics during Endometriosis Awareness Month in March when compared to February in 2020.

Regarding engagement, our results showed that post category had a substantial effect on post engagement, a correlation that had been previously observed [5]. Furthermore, we were able to display major differences in engagement between Instagram and Facebook. It is commonly known that engagement rates vary between social media platforms; content marketing providers assume "good" engagement rates in the range of 1% to 2% for Facebook, and 2% to 3% on Instagram [30].

As expected, we noticed higher user engagement on Instagram than on Facebook throughout all categories. When compared with the benchmarks described above, humoristic posts had exceptional engagement rates of more than 4% on Instagram, while many categories performed in the expected range of 2% to 3% (ie, "personal story," "inspiration and support," "awareness," "other," "education and research," "scientific inquiries," and "patient requests"). Instagram's superiority in this regard has been explained as being caused by its format, as users only see one post per screen, urging them to either engage or scroll past [30]. Different mechanisms apply for Facebook. Due to the fact that posts originating from pages are broadcast to a targeted minority of followers, who are more likely to engage, it has been described that Facebook has reduced the reach of such posts over the last years [30], possibly resulting in reduced engagement. Typically, humoristic or controversial content generates strong engagement, and other authors reported strong engagement with posts about oral contraceptives, intrauterine devices, or cancer [6].

Even though medical professionals only accounted for 15% of the Facebook pages and 5% of the Instagram accounts investigated, some of these physicians were able to connect and communicate with the community, rather than acting as passive bystanders [31]. As shown in our study, posts from health care professionals were able to generate high user engagement on Facebook. Although associated with a significant investment of mostly unpaid time as well as exposure, this additional information shared by medical professionals may lead to a more balanced and evidence-based presence of information.

Strengths

Several strengths of our study are noteworthy. Besides YouTube, which is mainly video based and was, therefore, not investigated, Facebook and Instagram are the two largest social media sites in the world, with around 50% of users being female. The approach of searching within Facebook's and Instagram's own search functions resembles a realistic user scenario when searching for endometriosis-related content on social media. In addition, a large number of posts that were shared with over 1 million followers were examined. By using a codebook, a profound and established method of categorizing posts was

applied, and the employment of two examiners, together with an arbitrator in case of disagreement, provided for additional objectivity. Lastly, analyzing data from two social media platforms with a partially overlapping user base allowed for identification of content-related differences between both services.

Limitations

Social media platforms are dynamic and ever-changing, limiting any attempt of mapping or categorizing their content. In this context, this study must be seen as a snapshot that provides insight but cannot adequately reflect the dynamic changes. Although we included two of the largest social media services, several other well-known social media sites, such as Twitter, LinkedIn, Snapchat, Pinterest, Reddit, and TikTok, have not been investigated. Therefore, no statement can be made about the distribution of topics on these services or on social media as a whole.

Facebook's and Instagram's proprietary search algorithms are noteworthy and have limited capabilities for end users. By presenting different search results depending on the user and location, certain pages (ie, accounts) that met the inclusion criteria may have been missed. On Instagram, sorting the results by the number of followers was not possible; this was addressed by screening the follower lists for further accounts with large follower numbers. On Facebook, sorting pages by the number of likes was possible at the time of data collection, a feature

that has since been disabled. This allowed for identification of the largest Facebook pages down to 5000 followers.

Regarding participation bias, it is arguable that only a certain subgroup of patients with endometriosis, or people interested in endometriosis in general, were actively sharing, discussing, and following endometriosis-related content. Assuming this is the case, this study was a content analysis study that used a subset of available posts on two social media platforms on a given date; therefore, participation or nonresponse bias may influence any study's findings in the first place, but it does not affect the veracity of this study's results.

Conclusions

Instagram and Facebook are being used intensively to share endometriosis-related information with a large number of users. The most common topics varied between platforms. Most posts offered inspiration or support, spread awareness about the disease, or covered personal issues. User engagement was higher on Instagram than on Facebook. Followers mostly engaged with posts with humoristic, awareness-generating, or personal content. Health care providers should be aware of the topics discussed online, as this may lead to an increased understanding of the needs and demands of digitally proficient patients with endometriosis. Future research should focus on topics that are trending as well as the scientific accuracy of social media content, possibly highlighting underlying drivers and interests of endometriosis-related content creators.

Authors' Contributions

JMM was responsible for conceptualization, methodology, investigation, preparation of the original draft of the manuscript, data visualization, and project administration. JMM, DRK, and LB were responsible for the formal analysis and data curation. All authors were responsible for reviewing and editing the manuscript. DRK was responsible for data validation. PI and GS were responsible for project supervision. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

References

1. Zondervan KT, Becker CM, Missmer SA. Endometriosis. N Engl J Med 2020 Mar 26;382(13):1244-1256. [doi: 10.1056/NEJMra1810764] [Medline: 32212520]

2. Shah DK, Moravek MB, Vahratian A, Dalton VK, Lebovic DI. Public perceptions of endometriosis: Perspectives from both genders. Acta Obstet Gynecol Scand 2010 May;89(5):646-650 [FREE Full text] [doi: 10.3109/00016341003657900] [Medline: 20235893]

3. Quibel A, Puscasiu L, Marpeau L, Roman H. General practitioners and the challenge of endometriosis screening and care: Results of a survey [Article in French]. Gynecol Obstet Fertil 2013 Jun;41(6):372-380. [doi: 10.1016/j.gyobfe.2012.02.024] [Medline: 22521982]

4. Hadfield R, Mardon H, Barlow D, Kennedy S. Delay in the diagnosis of endometriosis: A survey of women from the USA and the UK. Hum Reprod 1996 Apr;11(4):878-880. [doi: 10.1093/oxfordjournals.humrep.a019270] [Medline: 8671344]

5. Towne J, Suliman Y, Russell KA, Stuparich MA, Nahas S, Behbehani S. Health information in the era of social media: An analysis of the nature and accuracy of posts made by public Facebook pages for patients with endometriosis. J Minim Invasive Gynecol 2021 Sep;28(9):1637-1642. [doi: 10.1016/j.jmig.2021.02.005] [Medline: 33582381]

6. Carneiro MM, Farace BL, Ribeiro LSDC, Silverio RCCC, Moreira T, Filho ALDS, et al. Using social media to educate women and healthcare providers on endometriosis: Preliminary results. JBRA Assist Reprod 2020 Jan 30;24(1):9-12 [FREE Full text] [doi: 10.5935/1518-0557.20190048] [Medline: 31361436]

7. Auxier B, Anderson M. Social media use in 2021. Pew Research Center. 2021 Apr 07. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2021-09-14]

XSL•FO

RenderX

8.	Popular Instagram influencer topics worldwide 2020, by annual follower growth. Statista.: Statista Research Department; 2021 Oct 19. URL: https://www.statista.com/statistics/1117938/popular-instagram-influencer-topics-worldwide-follower-growth/ [accessed 2021-03-22]

9.	Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]

10.	Low JK, Manias E. Use of technology-based tools to support adolescents and young adults with chronic disease: Systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Jul 18;7(7):e12042 [FREE Full text] [doi: 10.2196/12042] [Medline: 31322129]

11.	Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]

12.	Nunes RH, Ferreira JB, de Freitas AS, Ramos FL. The effects of social media opinion leaders' recommendations on followers' intention to buy. Revista Brasileira de Gestão de Negócios 2018 Jan;20(1):57-73 [FREE Full text] [doi: 10.7819/rbgn.v20i1.3678]

13.	Bates BR, Romina S, Ahmed R, Hopson D. The effect of source credibility on consumers' perceptions of the quality of health information on the internet. Med Inform Internet Med 2006 Mar;31(1):45-52. [doi: 10.1080/14639230600552601] [Medline: 16754366]

14.	Hirsch M, Aggarwal S, Barker C, Davis CJ, Duffy JMN. Googling endometriosis: A systematic review of information available on the internet. Am J Obstet Gynecol 2017 Dec;216(5):451-458.e1. [doi: 10.1016/j.ajog.2016.11.1007] [Medline: 27840143]

15.	Burla L, Kalaitzopoulos DR, Metzler JM, Scheiner D, Imesch P. Popularity of endocrine endometriosis drugs and limited alternatives in the present and foreseeable future: A survey among 1420 affected women. Eur J Obstet Gynecol Reprod Biol 2021 Jul;262:232-238 [FREE Full text] [doi: 10.1016/j.ejogrb.2021.05.040] [Medline: 34091158]

16.	Canis M, Curinier S, Campagne-Loiseau S, Kaemerlen Rabischong AG, Rabischong B, Pouly JL, et al. Information to patients in endometriosis: We must stop the frightening machine! [Article in French]. Gynecol Obstet Fertil 2016 Nov;44(11):618-619. [doi: 10.1016/j.gyobfe.2016.09.009] [Medline: 27765429]

17.	Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: A systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]

18.	Mercier RJ, Senter K, Webster R, Henderson Riley A. Instagram users' experiences of miscarriage. Obstet Gynecol 2020 Jan;135(1):166-173. [doi: 10.1097/AOG.0000000000003621] [Medline: 31809440]

19.	Blakemore JK, Bayer AH, Smith MB, Grifo JA. Infertility influencers: An analysis of information and influence in the fertility webspace. J Assist Reprod Genet 2020 Jun;37(6):1371-1378 [FREE Full text] [doi: 10.1007/s10815-020-01799-2] [Medline: 32382959]

20.	Qin LA, El-Neemany D, Winkler H, Shalom D. #Urogyn: What's trending on Instagram? A cross-sectional observational study. Female Pelvic Med Reconstr Surg 2020 May;26(5):283-286. [doi: 10.1097/SPV.0000000000000869] [Medline: 32282524]

21.	van der Pijl MSG, Hollander MH, van der Linden T, Verweij R, Holten L, Kingma E, et al. Left powerless: A qualitative social media content analysis of the Dutch #breakthesilence campaign on negative and traumatic experiences of labour and birth. PLoS One 2020;15(5):e0233114 [FREE Full text] [doi: 10.1371/journal.pone.0233114] [Medline: 32396552]

22.	Gochi A, Coen N, Ge S, Stuparich M, Nahas S, Behbehani S. Instagram users' content on endometriosis–Does endometriosis awareness month make a difference? In: Proceedings of the American Society for Reproductive Medicine Scientific Congress & Expo. 2020 Presented at: American Society for Reproductive Medicine Scientific Congress & Expo; October 17-21, 2020; Virtual p. e213-e214 URL: https://www.fertstert.org/action/showPdf?pii=S0015-0282%2820%2931358-3 [doi: 10.1016/j.fertnstert.2020.08.604]

23.	Wilson S, Mogan S, Kaur K. Understanding the role of Facebook to support women with endometriosis: A Malaysian perspective. Int J Nurs Pract 2020 Aug;26(4):e12833. [doi: 10.1111/ijn.12833] [Medline: 32189457]

24.	Rayson M. What is a "Facebook engagement": Information on what is calculated in Facebook metrics. BuzzSumo Knowledge Base. URL: https://help.buzzsumo.com/en/articles/1964658-what-is-a-facebook-engagement [accessed 2020-07-15]

25.	Skalski PD, Neuendorf KA, Cajigas JA. Content analysis in the interactive media age. In: Neuendorf KA, editor. The Content Analysis Guidebook. 2nd edition. Thousand Oaks, CA: SAGE Publications, Inc; 2017.

26.	Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis. Int J Qual Methods 2017 Oct 02;16(1):1-13 [FREE Full text] [doi: 10.1177/1609406917733847]

27.	Sehl K. 6 ways to calculate engagement rate (free calculator). Hootsuite. URL: https://blog.hootsuite.com/calculate-engagement-rate/amp/ [accessed 2021-03-31]

28.	Randolph JJ. Free-marginal multirater kappa (multirater κfree): An alternative to Fleiss´fixed-marginal multirater kappa. In: Proceedings of the Joensuu Learning and Instruction Symposium. 2005 Presented at: Joensuu Learning and Instruction Symposium; October 14-15, 2005; Joensuu, Finland p. 1-20 URL: https://files.eric.ed.gov/fulltext/ED490661.pdf

29. Facebook tips: What's the difference between a Facebook page and group? Facebook. 2010 Feb 24. URL: https://es-la. facebook.com/notes/facebook/facebook-tips-whats-the-difference-between-a-facebook-page-and-group/324706977130/ [accessed 2021-05-09]
30. What is a good social media engagement rate? ContentCal. 2020 May 15. URL: https://www.contentcal.io/blog/what-is-a-good-social-media-engagement-rate/ [accessed 2021-04-30]
31. Quaas AM. Social media in ART-#power or #peril? J Assist Reprod Genet 2020 Jun;37(6):1311-1312 [FREE Full text] [doi: 10.1007/s10815-020-01831-5] [Medline: 32468332]

Abbreviations

ERR: engagement rate by reach

Original Paper

User Experiences With an SMS Text Messaging Program for Smoking Cessation: Qualitative Study

Alexandra Budenz[1], MA, DrPH; Kisha Coa[2], MPH, PhD; Emily Grenen[3], MSc; Brian Keefe[3], MA; Amy Sanders[3], MA; Kara P Wiseman[4], MPH, PhD; Maria Roditis[1], MPH, PhD

[1]Tobacco Control Research Branch, National Cancer Institute, Rockville, MD, United States

[2]Centers for Medicare and Medicaid Services, Woodlawn, MD, United States

[3]ICF, Fairfax, VA, United States

[4]Department of Public Health Sciences, University of Virginia, Charlottesville, VA, United States

Corresponding Author:
Emily Grenen, MSc
ICF
9300 Lee Hwy
Fairfax, VA, 22031
United States
Phone: 1 301 572 0369
Email: Emily.Grenen@icfnext.com

Abstract

Background: Mobile health strategies for smoking cessation (eg, SMS text messaging–based interventions) have been shown to be effective in helping smokers quit. However, further research is needed to better understand user experiences with these platforms.

Objective: This qualitative study aims to explore the experiences of real-world users of a publicly available smoking cessation program (SmokefreeTXT).

Methods: Semistructured phone interviews were conducted with 36 SmokefreeTXT users between March and July 2014. Of these 36 participants, 50% (18/36) of participants completed the SmokefreeTXT program (ie, did not opt out of the program before the 6- to 8-week completion period), and 50% (18/36) did not complete the program (ie, requested to opt out of the program before the completion period). Interview questions focused on smoking behaviors, quitting history, opinions on the program's content and structure, answering assessment questions, using keywords, reasons for opting out, and perceived usefulness of the program for quitting smoking. A thematic content analysis was conducted, with a focus on themes to increase program engagement and optimization.

Results: The findings highlighted features of the program that participants found beneficial, as well as some elements that showed opportunities for improvement to boost program retention and successful cessation. Specifically, most participants found the SmokefreeTXT program to be convenient and supportive of cessation; however, some found the messages to be repetitive and reported a desire for more flexibility based on their readiness to quit and cessation progress. We also found that program completion did not necessarily indicate successful smoking cessation and that program opt out, which might be interpreted as a less positive outcome, may occur because of successful cessation. Finally, several participants reported using SmokefreeTXT together with other evidence-based cessation methods or non–evidence-based strategies.

Conclusions: Qualitative interviews with real-world SmokefreeTXT users showed high program acceptability, engagement with program features, and perceived utility for smoking cessation. Our findings directly informed several program updates, such as adding an adaptive quit date feature and offering supplemental information on live support services for users who prefer human interaction during the cessation process. The study has implications for other digital tobacco cessation interventions and highlights important topics that warrant future research, such as the relationship between program engagement (eg, opt out and retention) and successful cessation.

XSL•FO

RenderX

Introduction

Background

Although cigarette smoking rates have declined over time, smoking remains the leading preventable cause of death in the United States [1]. In 2019, the prevalence of cigarette smoking among US adults was 13.7% [2]. Although evidence-based approaches, such as smoking cessation counseling and pharmacological cessation aids, can support cessation, they are underused [3,4]. Mobile health (mHealth) smoking cessation platforms (eg, mobile phones, smartphones, and tablets) can be beneficial to the cessation process, as they can reduce barriers (eg, time commitment and access limitations) associated with traditional modalities and have been found to facilitate smoking abstinence [5-9]. mHealth interventions also provide anonymity and real-time support, reaching users from the convenience of their mobile devices at any place or time of day. A type of mHealth intervention with demonstrated efficacy for smoking cessation is an SMS text messaging–based cessation program [5,10,11]. Given the popularity of these programs, pervasive ownership and use of mobile phones in the United States (97% of US adults report owning a mobile phone) [12], and opportunities to reach underserved populations with mHealth smoking cessation programs, it is important to continually optimize these programs.

Despite the demonstrated efficacy of SMS text message–based smoking cessation programs in quantitative evaluations, qualitative studies of user experiences with SMS text message cessation programs are scarce [13-15]. The few qualitative studies that exist on this topic have found high levels of acceptability toward SMS text message–based cessation programs, and participants have reported that they appreciate the convenience and emotional support that these programs offer. However, these studies have primarily been conducted as part of existing cessation trials or with specific populations such as women with pregnancies [13-16]. Moreover, prior studies have been conducted with participants who remained in the cessation SMS text messaging program under study for the entirety of the program, and the experiences of users who opted out of the program before completion were not investigated [13-15].

To address these gaps, we explore the experiences of real-world users of a publicly available smoking cessation program (SmokefreeTXT). Specifically, we conducted qualitative interviews with real-world users (ie, who were not part of an existing research study) of the National Cancer Institute's (NCI) publicly available SmokefreeTXT program to understand perceptions about program structure and content, engagement with specific program features, and perceived utility of the program for smoking cessation. We also investigate potential differences in program experience between users who completed the SmokefreeTXT program and those who opted out of the program before completion.

SmokefreeTXT Program

SmokefreeTXT is a free, publicly available, fully automated SMS text message–based smoking cessation program introduced in 2011 by the NCI [8]. The program was developed by a team of mobile technology specialists and clinical psychologists with expertise in tobacco cessation. In 2020, a total of 32,633 new users were enrolled in the SmokefreeTXT program designed for general adult smokers [7].

SmokefreeTXT comprises 6 to 8 weeks of SMS text messages that provide cessation motivation, tips on preparing to quit, advice on managing cravings, quit smoking facts, and recognition of cessation milestones. SmokefreeTXT also refers program users to Smokefree web resources for more detailed smoking cessation information, as well as to the National Network of Tobacco Cessation Quitlines and the NCI Cancer Information Service. Individuals can enroll in SmokefreeTXT through the Smokefree website or by texting a keyword to a short code (QUIT to 47848). At the time of enrollment, users are prompted to set a quit date within 2 weeks of enrollment. Depending on when users set their quit date, they can receive up to 2 weeks of preparation messages leading up to that date. Starting on the quit date, the 6-week intervention comprises 1 to 5 SMS text messages each day, including behavioral intervention and social support messages that target smoking cessation goals [7]. Messages also contain links to relevant pages on the Smokefree website and promote social support through Smokefree social media pages (eg, Facebook).

Throughout the program, users receive messages that assess their smoking status (eg, "Did you smoke today? Reply YES or NO"), craving level (eg, "What's your craving level? Reply with: HI, MED, or LOW"), and mood status (eg, "How is your mood? Reply GOOD, OK, or BAD"). Users can receive additional messages on demand by texting specific keywords that signify their needs (eg, *MOOD* if they are experiencing a negative mood, *CRAVE* if they have a craving, or *SLIP* if they have smoked a cigarette). Users may also opt out of the program at any time by texting the word *STOP* or reset their quit date at any time by texting the word *NEW* [7]. Completion of the program was defined as not opting out of the program before the conclusion of the 6- to 8-week intervention. See Multimedia Appendix 1 for examples of the program messages.

Methods

Recruitment and Participants

Between January and June 2014, all users who completed the SmokefreeTXT program or requested to opt out of the SmokefreeTXT program before program completion were eligible for the study. Program completers (defined as users enrolled in SmokefreeTXT for the duration of the 6- to 8-week program—depending on when they set their quit date—and not opting out before program completion) and noncompleters (defined as users who texted the word *STOP* to opt out of SmokefreeTXT before program completion) were sent an SMS text message inviting them to participate in a phone interview about the program. Interested participants were asked to complete a web-based screener that inquired about their age, race or ethnicity, sex, education, and location (state). Of the 285 individuals who completed the screener, 84.2% (240/285) were eligible to participate in the study (45/285, 15.8% were ineligible because of being aged <18 years or not including contact information on the screener). Eligible respondents were

contacted to schedule interviews. We purposively sampled an equal number of men and women as there are documented gender differences in smoking cessation [17]. We also purposefully sampled an equal number of completers and noncompleters, as prior studies of SmokefreeTXT showed differences in experiences with the program based on completion status [8,18]. The first 36 participants who met the recruitment goals (ie, half were men and half were women; half were program completers and half program were noncompleters) and were available to participate in the study were selected. The sample size was determined a priori because of timeline and funding logistics, widely cited literature on qualitative sample size [19-21], and project goals of maximizing variety and depth of the findings within each subgroup while also ensuring that the number of participants included met our purposive sampling criteria (equal numbers by gender and completion status).

Interviews were conducted by 2 experienced research associates with training in behavioral science and extensive experience conducting qualitative research. One of the researchers (Sondra Dietz) held a master's in public health, and the other (Bethany Tennant) held a doctorate in health education and behavior. Before the interview, participants were informed that the objective of the interview was to gather their feedback on SmokefreeTXT and that interviewers worked for ICF International Inc, a management consulting firm supporting the NCI. The interview questions focused on smoking behaviors, quitting history, opinions on the program's content and structure, answering assessment questions, using keywords (eg, texting CRAVE to receive additional messages related to smoking cravings), reasons for opting out, and perceived usefulness of the program for quitting smoking. The interview guide was reviewed by subject matter experts in mHealth interventions and smoking cessation but was not pilot-tested with adults who smoked. See Multimedia Appendix 2 for the interview guide. Most interviews lasted 30 to 40 minutes, and all transcripts were recorded and transcribed. Only the participants and members of the research team (ie, an interviewer and a notetaker [Sondra Dietz and Bethany Tennant]) were present on the calls. The interviewers read a consent script by phone and obtained verbal consent before the interview. Participants were compensated for their participation with a US $25 electronic gift card. After the compensation, participants were not contacted again (eg, to review transcripts or study findings).

Ethical Considerations

The study was approved by ICF's institutional review board (IRB), which holds Federalwide Assurance (FWA 00002349) from the HHS Office for Human Research Protections.

Data Analysis

The transcripts were analyzed using thematic analysis, an approach in which themes are extracted from qualitative data by identifying salient portions of the transcripts, codes are applied to salient text portions, and codes are extrapolated into larger themes according to their relationship with the study topics of interest [22]. This process mainly relied on deductive coding [23], which was informed by themes that could potentially guide program optimization, such as feedback on program structure and content, but also allowed flexibility for

additional codes that emerged. The data analysis was conducted by 3 ICF staff members and 1 NCI staff member, all of whom were trained in either health behavior, public health, or both, and were experienced qualitative researchers. To develop the initial codes, 4 transcripts were repeatedly reviewed, notated with analytic memos throughout the coding process, and analyzed to devise a list of codes. The coders then met to discuss the codes and any discrepancies to reach a consensus, after which point, a preliminary codebook was formed [24]. Next, 2 transcripts were coded using the preliminary codebook. The codebook was further refined based on discussions between the coders to clarify definitions and remove duplicative codes. Using the final codebook, all 36 transcripts were coded in NVivo (QSR International). This process was iterative, and if new codes emerged during the full analysis, the coders convened to discuss the code, gain consensus, and update the codebook and previously coded transcripts accordingly. After all transcripts were coded, groups of codes were categorized and interpreted as larger themes [23]. Themes were further refined through a rereview of the transcripts using negative case analysis, wherein the coders discussed findings from *outlier* participants whose reports conflicted with most of the sample [25]. For example, most participants reported opting out of the SmokefreeTXT program as they began smoking again or successfully quit smoking; however, a small number of participants reported opting out for other reasons (eg, technology issues), causing the initial hypothesis that opt out was driven only by smoking status to be reconsidered. Although coders looked for differences in responses between men and women and completers and noncompleters, no substantive differences were found. Thus, results are reported for all participants combined, except for the results from questions only asked to noncompleters.

Results

Overview

A total of 36 adults, of whom 18 (50%) had completed the program (*completers*) and 18 (50%) had not (*noncompleters*), participated in semistructured, in-depth, individual telephone interviews conducted between March 2014 and July 2014. Most participants had some college education or higher (24/36, 67%), were White (29/36, 81%), and were from the Southern or Northeastern United States (24/36, 67%). The participants' average age was 36 years (SD 13.5; range 18-60 years; Table 1).

At the time of the interview, participants were asked about their cigarette smoking status, and 53% (19/36) of participants reported that they were no longer smoking (13/18, 72% of completers and 6/18, 33% of noncompleters), whereas 47% (17/36) reported that they were current smokers (5/18, 28% completers, and 12/18, 67% noncompleters). Half of the noncompleters who reported smoking at the time of the interview (6/12, 50%) smoked a pack of cigarettes or more per day, whereas only one of the completers who reported smoking at the time of the interview smoked a pack or more per day. Some participants also reported the use of other tobacco products (eg, cigars, e-cigarettes, hookah, and smokeless tobacco) during the interviews. Approximately 31% (11/36) of participants

reported current use of other tobacco products at the time of the interview (3/18, 17% completers and 8/18, 44% noncompleters). See Table 1 for tobacco use behaviors in the sample.

Several primary themes emerged during the thematic analysis of participant interviews. Participants discussed their experiences with the SmokefreeTXT program and perceptions of message content, timing, and frequency. Participants also discussed their use of program engagement features, such as on-demand keywords (eg, *CRAVE*) and assessment questions (eg, *How is your mood today?*). Program noncompleters discussed the reasons for opting out of the program before program completion. Participants also discussed the role of SmokefreeTXT in their quitting process and reported the concurrent use of other cessation strategies with SmokefreeTXT.

Table 1. SmokefreeTXT user participants' demographics and tobacco use behaviors (N=36)[a].

Characteristics	Total sample	Completers (n=18)	Noncompleters (n=18)
Gender, n (%)			
Male	18 (50)	9 (50)	9 (50)
Female	18 (50)	9 (50)	9 (50)
Age (years), mean (SD; range)	35.9 (13.5; 18-60)	39.3 (12.4; 19-60)	32.5 (13.7; 18-55)
Race or ethnicity, n (%)			
White	29 (81)	14 (78)	15 (83)
Black	5 (14)	3 (17)	2 (11)
Hispanic	1 (3)	0 (0)	1 (6)
Asian	1 (3)	1 (6)	0 (0)
Education, n (%)			
Less than high school	1 (3)	0 (0)	1 (6)
High school or equivalent	11 (31)	4 (22)	7 (39)
Some college	12 (33)	8 (44)	4 (22)
Graduated college	10 (28)	4 (22)	6 (33)
Postcollege education	2 (6)	2 (11)	0 (0)
Geographic region, n (%)			
Northeast	10 (28)	7 (39)	5 (28)
South	14 (39)	7 (39)	5 (28)
Midwest	6 (17)	3 (17)	4 (22)
West	6 (17)	1 (6)	4 (22)
Smoking status at time of interview[b], n (%)			
Smoker	17 (47)	5 (28)	12 (67)
Ex-smoker[c]	19 (53)	13 (72)	6 (33)
Smoking heaviness at time of interview[d] (n=17), n (%)			
<1 (pack per day)	10 (59)	4 (80)	6 (50)
≥1 (pack per day)	7 (41)	1 (20)	6 (50)
Other tobacco use at time of interview, n (%)			
Other tobacco user	11 (31)	3 (17)	8 (44)
Not other tobacco user	25 (69)	15 (83)	10 (56)

[a]Owing to rounding, some percentages may not add up to 100%.

[b]Asked of each participant during the interview.

[c]Defined as participants who reported that they did not smoke cigarettes; may include people who reported using other tobacco products.

[d]Asked of each participant who reported smoking at the time of the interview (n=17).

XSL•FO
RenderX

Participant Experiences With SmokefreeTXT

Convenience of a Mobile-Based Medium

Participants responded positively to the mobile phone–based delivery of SmokefreeTXT, noting its convenience. Some reported that they appreciated the SMS text messaging format as they had already spent a great deal of time on their phones:

> I like the whole idea because I'm always on my phone and so, you know, consistent text messages were, you know, great reminders of the fact that I am quitting and that I want to quit and I need to quit. So, I liked it. [program completer, female, cigar smoker]

Others similarly noted they could quickly read a text, which was preferable to engaging with more time-consuming cessation resources:

> I like the text messages because it pops up during the day, and the job that I do, that works out better, because I check my text messages every so often. And I can't never answer the phone, and I don't have time to search the web. But I've got time to get a really quick text. [program noncompleter, female, smoker]

Accountability and Social Support

Several participants reported that they liked having the program *check up* on them, especially when messages aligned with cravings:

> ...It's just an excellent reminder for me, especially if I happen to be thinking about doing that [smoking]...all of a sudden...I get this text...it just was an excellent reminder. [program completer, male, smoker]

Approximately half of the participants said they felt that SmokefreeTXT was supportive, and one of the participants said it was "like my pocket buddy" (program completer, male, ex-smoker). Some reported that they felt especially supported by SmokefreeTXT as they lacked social support from family and friends or were hesitant to share their quit attempts with others, as they felt judged by previous failed quit attempts:

> ...None of my friends like here are like smokers, so I can't tell them like "oh my god, I'm craving a cigarette so bad right now"...I guess this might sound weird, but like I could text someone who like understands... [program completer, female, ex-smoker]

Automation and Human Interaction

Many participants reported that they disliked that the program sent automated messages and did not include human interaction, reporting that the messages were repetitive, *too automated*, and like *cookie-cutter lines*:

> Sometimes, I'd get the same message, and it wouldn't be the same messages in a row, but it seemed like sometimes, it was pretty generic...It'd be the same sort of three messages. [program noncompleter, male, ex-smoker]

A participant reported strongly disliking the program, explaining that he did not feel that an automated program that did not directly provide live support gave enough assistance to promote successful cessation:

> Because it was talking to a machine...Some of it was just repetitive. There was no real emotional support there at all. [program completer, male, smoker]

Message Tone and Content

The participants had varying perspectives on the preferred message tone. Approximately half of the participants reported that they preferred the positive tone of the program, with messages focusing on the benefits of quitting rather than the negative aspects of smoking. Participants reported that a supportive tone was particularly beneficial when they were facing struggles during their quit attempts:

> I liked that, like if you did slip, it wasn't like oh, my God, how dare you. It would be like...do you want to start over again? Because like that's happened to me before. It was good. [program completer, female, ex-smoker]

However, some participants wanted a *tougher* tone to encourage accountability:

> Give me the more serious stuff. Don't tell me it's okay to have a slip, because honey, I'll have slips all day then...I need some really butt-kicking stuff. [program completer, female, smoker]

Some participants reportedly preferred a mix of message tones depending on their mood or how confident they felt about their quit attempt:

> I think it would be half and half. If you leave work, you might need a more supportive message if you had a bad day. [program noncompleter, male, ex-smoker]

Finally, some participants reported that the content and tone of the messages may have been an impediment to cessation. A few said that the messages triggered smoking cravings, and some noted that certain messages seemed to *permit* smoking by framing quit attempt failures positively:

> ...One that said, even if you fail, if you smoke, don't give up...like it was ok...I slipped once. I was like, no, don't tell me that it's ok to slip once. [program noncompleter, female, ex-smoker]

Message Frequency, Timing, and Duration

Most participants reported that the number of daily messages received in the program was appropriate; however, some mentioned that they would have preferred either more or fewer messages per day. In addition, several participants suggested that the timing of messages could align better with their personal schedules or typical smoking times:

> If you could set the timing, that would be a plus...I tend to smoke early in the morning and late at night...if you can set the frequency on when you do it, that would be an excellent plus because you're going to get that reminder at the times that you

actually may be smoking. [program completer, male, smoker]

Some participants also wanted to be able to customize message frequency or program duration so that the program could be abbreviated if the user successfully quit or extended if the user was struggling to quit:

> *...I wanted more messages...because I wasn't doing so good in any of my quits. Whereas, like this time next week, you know, maybe I am...doing awesome and it will be the perfect timing...Like maybe there could be some type of...where it says "crave" or "doing good"...just maybe "having a rough time"...and that's a trigger to the system to say..."extend her out for three more weeks."* [program completer, female, smoker]

A few participants noted that having an adaptive quit date (eg, 1 in the past and 1 that could be postponed) would make the program better suited to fluctuating levels of readiness to quit.

Program Engagement Features

Program Keywords

Most participants noted that they liked using the keyword feature when they were struggling with their quit attempt:

> *It responds quickly when you text "crave." I like the instant response when you tell it that you are having a problem. It gives me ideas of how to stay strong.* [program noncompleter, female, smoker]

However, some participants did not use the keywords as they did not remember that this feature was available to them.

Assessment Questions

Most participants reported that they liked receiving and responding to the assessment questions. Participants reported a variety of reasons for liking assessments, including their interactivity, accountability, and utility for smoking trigger recognition:

> *...if I would get the [assessment] message during a particular time of day, it would help then...I would stop and think about the things that were going on around me at that time, and then based on what I replied to the message and what my craving level was...I could say...these are the things that are making me crave.* [program completer, male, ex-smoker]

However, many participants reported barriers in responding to assessments, including not having time to respond, not recalling the messages, or feeling guilty for having smoked or relapsed:

> *I'm smoking, and yet, it thinks that I've been quit for five days. Don't tell. So, I'm not responding to it, but that's because I know I slipped up. But at the same time, I still want the messages to come through.* [program completer, female, smoker]

Program Opt Out (Noncompletion)

Smoking Status and Opt Out

Among program noncompleters, more than half reported opting out because they were not successful in quitting smoking. A few participants expressed guilt about continuing to use the program while they were still smoking:

> *Why did I opt out? Well, I started smoking again...It just would make me feel like I was lying, you know? Even if I typed in "slip," when I throw back a pack and a half a day, that's not a slip. That's a habit again.* [program noncompleter, male, smoker]

Some noncompleters reported opting out as they had successfully quit smoking, and one of these participants reported opting out as the messages were becoming a trigger to smoke after he had quit.

Program Factors Influencing Opt Out

Some noncompleters reported that they opted out as a result of the volume and content of the SMS text messages. Some reported that the number of texts received per day contributed to opt out, either reporting that the number of texts per day was too high or that the decrease in messages over time was a factor in opting out:

> *They kind of slowed down the texts, and...it kind of made me feel that I didn't need it anymore...The messages weren't as strong.* [program noncompleter, female, ex-smoker]

Some noncompleters also reported that they opted out because the messages on their own were not enough of a motivator to quit:

> *I'd read the text...and then soon as I would close it out light a cigarette within two minutes...I don't think it did a trigger...It just didn't do anything to prevent it...it didn't really help me to say, "okay, I'm not going to."* [program noncompleter, male, smoker]

Other Factors Influencing Opt Out

Along with smoking status and SMS text message–related factors that contributed to opt out, a few participants reported that they had technical issues with their mobile phones that contributed to opt out. One of the participants reported that he no longer needed the program because he had enough social support at the time.

The Role of SmokefreeTXT in the Quitting Process

Several participants reported that they were still smoking at the time they enrolled in SmokefreeTXT, whereas fewer said they had started their quit attempt before enrollment. Several participants reported that they joined SmokefreeTXT during one of the multiple quit attempts, and some reported that they were enrolled in the program multiple times, re-enrolling in the program after completion or opt out:

> *...I was like, you know, I really want to quit. So I thought like maybe this [SmokefreeTXT] will work. And I tried it the first time and I quit for a while and...I had to restart it because I started smoking*

again. And then I was like I quit for a while and then like things just got crazy and I started smoking and then I was like "I'm going to text the [SmokefreeTXT] number again and it's going to be the last time I do it." [program completer, female, ex-smoker]

Approximately half of the participants reported that they felt that SmokefreeTXT had a positive impact on their quitting process, calling the program *helpful* and reporting that it made them feel more confident about quitting. Several participants who were smoking at the time of the study reported that, although they did not fully quit while enrolled in the program, the program helped them quit temporarily or reduce their smoking (fewer cigarettes per day and smoking less frequently).

Other Cessation Support in Conjunction With SmokefreeTXT

Half of the participants reported that although SmokefreeTXT was a useful component of their quit plan, they also relied on other strategies. Notably, several participants reported using strategies that were not evidence-based, such as switching to smokeless tobacco or using hypnosis rather than evidence-based methods such as over-the-counter nicotine replacement therapy (NRT) or prescription cessation medications. Although there are no reports on the implications for SmokefreeTXT on the effectiveness of or adherence to NRT, a couple of participants reported they were motivated to use NRT while they were enrolled in the SmokefreeTXT program, and a few participants reported that they would like SmokefreeTXT to facilitate access to NRT, such as by providing coupons. Several participants also reported using other mHealth cessation resources while enrolled in SmokefreeTXT, such as cessation smartphone apps and social media focused on cessation support (eg, Smokefree Women Facebook page and web-based smoking cessation forum). Some participants who reported using cessation smartphone apps reported that they would like features of cessation apps to be incorporated into SmokefreeTXT, including tracking money saved by not smoking and smoking milestones (number of days since quitting).

Discussion

Principal Findings

Interviews with SmokefreeTXT users showed high program acceptability, engagement with program features, and perceived utility for smoking cessation. Participants provided feedback on program content and functionality, which informed changes to the program. Moreover, the findings provide insights into the complex relationship between program engagement and cessation, which warrants future research.

This qualitative study adds to the limited body of literature that assesses user experiences of smoking cessation SMS text messaging programs. A unique component of this study is its focus on the experiences and perceptions of real-world users of the publicly available SmokefreeTXT smoking cessation SMS text messaging program, which, to our knowledge, has not been explored by other studies of cessation SMS text messaging programs [13-16]. The findings highlighted features of the program that participants found beneficial, as well as some

elements that showed opportunities for improvement to boost program retention and successful cessation. We also found that program completion does not necessarily indicate successful smoking cessation and that program opt out, which may be interpreted as a less positive outcome [18], may occur because of successful cessation. Finally, several participants reported using SmokefreeTXT together with other evidence-based cessation methods or non–evidence-based strategies.

As has been found in prior qualitative studies, participants reported that SmokefreeTXT was easier to use than other live support services that require a more significant time commitment [13,14]. They also noted that the anonymity of the platform provided a source of discrete support without perceived judgment or stigma from family or friends [14]. However, similar to what has been found previously, several participants disliked the automated nature of the program, the absence of human interaction, and what they perceived as generic-sounding messages [14,15].

In terms of program structure, many participants reported that they would like the ability to customize message timing, program duration, and quit date settings, which has been reported in previous studies [13,16]. This underscores the need for mobile-based programs to take into consideration that cessation is a highly individualized and often nonlinear process, which may be optimally supported by programs offering flexible structure and functionality.

Although participants liked that the keyword feature allowed them to receive on-demand support and that the smoking assessment questions provided accountability, in some instances, participants reported that assessments triggered feelings of guilt if they had relapsed, which mirrors previous research [13]. In addition, some participants felt that they could not respond accurately to assessment questions (eg, a question about *slips* may not feel relevant if a user had relapsed). Therefore, as quitting smoking is rarely a linear process, it is important to consider other metrics of program success in addition to abstinence [26]. For example, some participants in our study reported that the SmokefreeTXT program helped them quit temporarily, and participants commonly reported making multiple quit attempts, both of which have been defined as metrics of success in other cessation SMS text messaging studies [13,15]. Participants also noted a reduction in their smoking. Given that many study participants entered the program during one of several quit attempts and that several participants enrolled in the program multiple times, further research is needed to better understand how to meet the needs of smokers who may engage in several quit attempts before being successful. For users who are not yet ready or able to quit, examining intermediate outcomes (potentially through responses to assessment questions), such as the number of quit attempts or cutting back on smoking, may provide valuable information to best support users during different stages of the cessation process.

To our knowledge, this is one of the first qualitative studies of cessation text program users that has explored the factors influencing program opt out. Two existing studies compared the experiences of users who either quit smoking or did not quit

XSL·FO
RenderX

smoking while using a cessation SMS text messaging program [13,15]. As participants of those studies were required to complete the program as part of their participation in an existing clinical trial, the studies were unable to examine the characteristics and experiences of users who would have otherwise opted out of the program. Most noncompleters reported opting out of SmokefreeTXT because of smoking relapse. However, not all participants who opted out did so as they had resumed smoking. Some opted out as they had successfully quit smoking and felt that they no longer needed the program, which suggests a need to reconceptualize the implications of program opt out [27]. Therefore, although certain factors may make opt out more likely, this qualitative work underscores that opt out is more nuanced than previous research suggests. In other words, opt out may not be definitively used as an indicator of either continued smoking or abstinence; rather, it is best seen as an indication that the user no longer finds it helpful to receive messages from an SMS text messaging program.

These findings also suggest that strategies to re-engage users that opt out should not be a one-size-fits-all approach. Although smoking relapse and cessation were the primary drivers of opt out, a few participants cited the number of messages as a reason for opting out, and others reported opting out as they did not find the messages sufficiently motivating to help them quit smoking. This reinforces the need to consider the message volume and tone when developing an automated program [13,14]. Opt out may have also been influenced by the level of smoking dependence and other tobacco use in the noncompleters group, as noncompleters tended to be heavy smokers, and many were poly-tobacco users.

Finally, although most participants reported that SmokefreeTXT was a useful component of a quit plan, approximately half relied on other cessation strategies as a complement to SmokefreeTXT. Notably, many of the strategies used by study participants were not evidence-based, and some participants reported supplementing their use of SmokefreeTXT by switching from cigarettes to other tobacco products, exposing them to further health risks associated with tobacco use. Therefore, it is important to consider how SMS text message programs can be used to encourage participants to both find evidence-based support (eg, NRT and prescription cessation medication) and work synergistically with these other supports.

SmokefreeTXT Optimizations

The results of this study have informed several optimizations to the program. For example, SmokefreeTXT has added messages that provide supplemental information on live support services and has optimized the SMS text message library to ensure that SMS text messages are more varied and less repetitive. Since conducting this study, the SmokefreeTXT program has also been changed to allow users to have a more flexible quit date at the start of the program. Users can set a quit date in the past if they begin their quit attempt before starting the SmokefreeTXT program, and users are also sent a message the day before their scheduled quit date and asked to text *ready* if they would like to keep the quit date they set or *not ready* to postpone the start of their quit attempt and set a new date.

Although we conducted qualitative interviews with users of the SmokefreeTXT program, several findings may also be applicable to the development and optimization of other cessation SMS text messaging programs. First, it may be important to allow users to customize the number of SMS text messages that they receive in the program. Cessation SMS text messaging programs may also be used with other cessation strategies; thus, it may be beneficial for SMS text message content to include messages about evidence-based cessation strategies. Relatedly, live support (eg, tobacco quitlines) is a cessation resource that can be recommended within SMS text message programs to benefit users who prefer human interaction as a component of their cessation process. Finally, program retention versus opt out may have limitations in the ability to predict the cessation *success* of users who opt out or remain in the program.

Limitations

This study had several limitations. First, we conducted in-depth interviews with a small convenience sample of SmokefreeTXT users who were mostly young, White, and had at least some college education. Thus, the results may not be generalizable to a wider population of smokers in the United States. Most current users of SmokefreeTXT are White (11,383/18,510, 61.50%), which demonstrates the need to consider how the program can better reach other demographic groups. In 2016, the SmokefreeTXT program began collecting race and ethnicity data as part of program opt in to track the demographics of the program over time, and future research recruitment efforts may be informed by these data.

SmokefreeTXT users who agreed to participate in the study may have been highly motivated to provide feedback on the program, whether positive or negative, which may have influenced the results. In addition, smoking status was not biochemically verified and relied on the self-reports of the users during the interviews. It is possible that some users stated that they had quit smoking to be more socially acceptable to the interviewer.

This study was conducted in 2014; since then, both mobile phone technology and the SmokefreeTXT program have evolved. However, although there have been changes to message content and quit date flexibility over time, the overarching structure and features of the SmokefreeTXT program (ie, message timing, keywords, and assessment questions) have remained consistent. Furthermore, the results still provide unique data from users who were not part of an intervention trial, and the SmokefreeTXT program continues to be widely used. Thus, the results can still provide actionable information for those developing new or adapting established cessation SMS text messaging interventions. Given the increasing importance of mHealth as an avenue for cessation, data from this study are still relevant today. In addition, some of the findings presented here may have new relevance as improvements to technology enable some of the recommended optimizations, such as the integration of SMS text messages with other digital resources; for example, quitlines or chat applications.

Conclusions

Qualitative interviews with real-world SmokefreeTXT users showed high program acceptability, engagement with program features, and perceived utility for smoking cessation. Our findings demonstrate the importance of allowing customization to the frequency of SMS text messages within cessation text programs, the limitations of measuring cessation *success* through program opt out versus retention, and the importance of recommending evidence-based cessation resources (including live support via quitlines for users who prefer human interaction during the cessation process) to be used with smoking cessation text programs.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example SmokefreeTXT program messages.
[DOCX File , 21 KB - formative_v6i3e32342_app1.docx]

Multimedia Appendix 2
Example Moderator Guide.
[DOCX File , 25 KB - formative_v6i3e32342_app2.docx]

References

1. Smoking cessation: a report of the surgeon general. Centers for disease control and prevention. 2020. URL: https://www.cdc.gov/tobacco/data_statistics/sgr/2020-smoking-cessation/index.html [accessed 2021-03-05]

2. Cornelius ME, Wang TW, Jamal A, Loretan CG, Neff LJ. Tobacco product use among adults - United States, 2019. MMWR Morb Mortal Wkly Rep 2020;69(46):1736-1742 [FREE Full text] [doi: 10.15585/mmwr.mm6946a4] [Medline: 33211681]

3. Clinical Practice Guideline Treating Tobacco Use and Dependence 2008 Update Panel, Liaisons,Staff. A clinical practice guideline for treating tobacco use and dependence: 2008 update. A U.S. Public Health Service report. Am J Prev Med 2008;35(2):158-176 [FREE Full text] [doi: 10.1016/j.amepre.2008.04.009] [Medline: 18617085]

4. Babb S, Malarcher A, Schauer G, Asman K, Jamal A. Quitting smoking among adults - United States, 2000-2015. MMWR Morb Mortal Wkly Rep 2017;65(52):1457-1464 [FREE Full text] [doi: 10.15585/mmwr.mm6552a1] [Medline: 28056007]

5. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016;4(4):CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]

6. Abroms LC, Boal AL, Simmens SJ, Mendel JA, Windsor RA. A randomized trial of Text2Quit: a text messaging program for smoking cessation. Am J Prev Med 2014;47(3):242-250 [FREE Full text] [doi: 10.1016/j.amepre.2014.04.010] [Medline: 24913220]

7. Prutzman YM, Wiseman KP, Grady MA, Budenz A, Grenen EG, Vercammen LK, et al. Using digital technologies to reach tobacco users who want to quit: evidence from the national cancer institute's smokefree.gov initiative. Am J Prev Med 2021;60(3 Suppl 2):S172-S184 [FREE Full text] [doi: 10.1016/j.amepre.2020.08.008] [Medline: 33663705]

8. Augustson E, Cole-Lewis H, Sanders A, Schwarz M, Geng Y, Coa K, et al. Analysing user-reported data for enhancement of SmokefreeTXT: a national text message smoking cessation intervention. Tob Control 2017;26(6):683-689. [doi: 10.1136/tobaccocontrol-2016-052945] [Medline: 27852892]

9. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]

10. Scott-Sheldon LA, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text messaging-based interventions for smoking cessation: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016;4(2):e49 [FREE Full text] [doi: 10.2196/mhealth.5436] [Medline: 27207211]

11. Spohr SA, Nandy R, Gandhiraj D, Vemulapalli A, Anne S, Walters ST. Efficacy of SMS text message interventions for smoking cessation: a meta-analysis. J Subst Abuse Treat 2015;56:1-10. [doi: 10.1016/j.jsat.2015.01.011] [Medline: 25720333]

XSL•FO
RenderX

12.   Mobile fact sheet. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-03-05]

13.   Douglas N, Free C. 'Someone batting in my corner': experiences of smoking-cessation support via text message. Br J Gen Pract 2013;63(616):e768-e776 [FREE Full text] [doi: 10.3399/bjgp13X674459] [Medline: 24267860]

14.   Naughton F, Jamison J, Sutton S. Attitudes towards SMS text message smoking cessation support: a qualitative study of pregnant smokers. Health Educ Res 2013;28(5):911-922. [doi: 10.1093/her/cyt057] [Medline: 23640985]

15.   Sloan M, Hopewell S, Coleman T, Cooper S, Naughton F. Smoking cessation support by text message during pregnancy: a qualitative study of views and experiences of the MiQuit intervention. Nicotine Tob Res 2017;19(5):572-577 [FREE Full text] [doi: 10.1093/ntr/ntw241] [Medline: 28403457]

16.   Bock BC, Heron KE, Jennings EG, Magee JC, Morrow KM. User preferences for a text message-based smoking cessation intervention. Health Educ Behav 2013;40(2):152-159. [doi: 10.1177/1090198112463020] [Medline: 23086555]

17.   Smith PH, Bessette AJ, Weinberger AH, Sheffer CE, McKee SA. Sex/gender differences in smoking cessation: a review. Prev Med 2016;92:135-140 [FREE Full text] [doi: 10.1016/j.ypmed.2016.07.013] [Medline: 27471021]

18.   Coa KI, Wiseman K, Higgins B, Augustson E. Associations between engagement and outcomes in the SmokefreeTXT program: a growth mixture modeling analysis. Nicotine Tob Res 2019;21(5):663-669 [FREE Full text] [doi: 10.1093/ntr/nty073] [Medline: 29668984]

19.   Guest G, Bunce A, Johnson L. How many interviews are enough?: an experiment with data saturation and variability. Field Methods 2006;18(1):59-82 [FREE Full text] [doi: 10.1177/1525822x05279903]

20.   Creswell JW. Qualitative inquiry and research design: choosing among five traditions. Thousand Oaks, CA: Sage Publications; 1998.

21.   Morse JM. Determining sample size. Qual Health Res 2000;10(1):3-5 [FREE Full text] [doi: 10.1177/104973200129118183]

22.   Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017;16(1):160940691773384 [FREE Full text] [doi: 10.1177/1609406917733847]

23.   Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]

24.   Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods 2006;5(1):80-92 [FREE Full text] [doi: 10.1177/160940690600500107]

25.   Creswell JW. Qualitative inquiry & research design: choosing among five approaches. 3rd edition. Thousand Oaks, CA: Sage Publications; 2013.

26.   Tan J, Zhao L, Chen H. A meta-analysis of the effectiveness of gradual versus abrupt smoking cessation. Tob Induc Dis 2019;17:09 [FREE Full text] [doi: 10.18332/tid/100557] [Medline: 31582921]

27.   Christofferson DE, Hertzberg JS, Beckham JC, Dennis PA, Hamlett-Berry K. Engagement and abstinence among users of a smoking cessation text message program for veterans. Addict Behav 2016;62:47-53 [FREE Full text] [doi: 10.1016/j.addbeh.2016.06.016] [Medline: 27318948]

Abbreviations

mHealth: mobile health
NCI: National Cancer Institute
NRT: nicotine replacement therapy

XSL•FO
RenderX

<u>Original Paper</u>

Improving Outcomes Through Personalized Recommendations in a Remote Diabetes Monitoring Program: Observational Study

Sowmya Kamath[1], PhD; Karthik Kappaganthu[1], PhD; Stefanie Painter[1], DHEd; Anmol Madan[1], PhD

Teladoc Health, Purchase, NY, United States

Corresponding Author:
Stefanie Painter, DHEd
Teladoc Health
2 Manhattanville Rd
Purchase, NY, 10577
United States
Phone: 1 3123307236
Email: spainter@teladoc.com

Abstract

Background: Diabetes management is complex, and program personalization has been identified to enhance engagement and clinical outcomes in diabetes management programs. However, 50% of individuals living with diabetes are unable to achieve glycemic control, presenting a gap in the delivery of self-management education and behavior change. Machine learning and recommender systems, which have been used within the health care setting, could be a feasible application for diabetes management programs to provide a personalized user experience and improve user engagement and outcomes.

Objective: This study aims to evaluate machine learning models using member-level engagements to predict improvement in estimated A_{1c} and develop personalized action recommendations within a remote diabetes monitoring program to improve clinical outcomes.

Methods: A retrospective study of Livongo for Diabetes member engagement data was analyzed within five action categories (interacting with a coach, reading education content, self-monitoring blood glucose level, tracking physical activity, and monitoring nutrition) to build a member-level model to predict if a specific type and level of engagement could lead to improved estimated A_{1c} for members with type 2 diabetes. Engagement and improvement in estimated A_{1c} can be correlated; therefore, the doubly robust learning method was used to model the heterogeneous treatment effect of action engagement on improvements in estimated A_{1c}.

Results: The treatment effect was successfully computed within the five action categories on estimated A_{1c} reduction for each member. Results show interaction with coaches and self-monitoring blood glucose levels were the actions that resulted in the highest average decrease in estimated A_{1c} (1.7% and 1.4%, respectively) and were the most recommended actions for 54% of the population. However, these were found to not be the optimal interventions for all members; 46% of members were predicted to have better outcomes with one of the other three interventions. Members who engaged with their recommended actions had on average a 0.8% larger reduction in estimated A_{1c} than those who did not engage in recommended actions within the first 3 months of the program.

Conclusions: Personalized action recommendations using heterogeneous treatment effects to compute the impact of member actions can reduce estimated A_{1c} and be a valuable tool for diabetes management programs in encouraging members toward actions to improve clinical outcomes.

(JMIR Form Res 2022;6(3):e33329) doi:10.2196/33329

Introduction

Diabetes is a chronic progressive disease affecting 34 million Americans with 1.5 million newly diagnosed each year [1,2]. Individuals living with diabetes are at greater risk of health complications including increased hospitalizations that result in 1 of every 4 health care dollars spent on diabetes-related care in the United States [3]. An essential factor in successfully living with diabetes is effective self-management, which has shown to improve glycemic control and reduce hospital admissions and the overall lifetime cost of health care [4].

Diabetes self-management efficacy and improved glycemic control is supported by programs that offer education, coaching, glucose monitoring, and physical activity [5-7]. Diabetes management programs have shown to be as or more effective than usual care in providing a significant reduction in hemoglobin A_{1c} (HbA_{1c}) [8-10]. Additionally, structured self-monitoring of blood glucose (SMBG) has been observed to improve glycemic variability and provide greater self-efficacy in management by helping an individual understand lifestyle behaviors' impact on blood glucose (BG) values over time [2,11]. However, while advances in diabetes treatment options, diabetes management programs, new technologies to support self-management, and the rise in digital health are rushing the market, half of individuals with diabetes have an HbA_{1c} value of 7.0% or higher and struggle to obtain consistent glycemic control [1]. This alludes to potential gaps in self-management programs, technology, and delivery to the individual [12].

Personalization has been identified as a key tool in digital health to enhance user engagement for improved outcomes, which is often a missing factor in the development of diabetes digital health programs [12]. Self-management best practices and user preference must be taken into consideration to effectively provide a personalized experience within a diabetes management program. This study has proposed and analyzed the feasibility of using heterogeneous treatment effect models for personalizing action recommendations within a digital remote diabetes monitoring program (RDMP).

Methods

Livongo for Diabetes

Livongo for Diabetes is an RDMP focused on empowering members with education and tools to self-manage their diabetes through mobile technology. The program offers members a cellular-enabled, two-way messaging device that measures BG and delivers personalized insights into their glycemic management; free unlimited BG test strips; real-time support from diabetes response specialists 24 hours a day, 7 days a week, 365 days a year; and access to certified diabetes care and education specialists (CDCESs) for support and goal setting.

Livongo members' glucose meter use was captured remotely through the cellular-enabled device. Members also had access to a mobile phone app that tracked historical SMBG readings and provided reminders for SMBG checking, physical activity, and food log tracking; asynchronous chat with coaches; ability to schedule private coaching sessions with CDCES; educational content for diabetes self-management; and allowed members to send historical reports of SMBG readings to care providers, family members, and friends.

Study Design

A retrospective feasibility study was conducted to compute heterogeneous treatment effect for five different action categories in the reduction of estimated A_{1c} for members enrolled in Livongo for Diabetes with type 2 diabetes and to identify which actions could be most effective for each member. Within each action category, members were classified into a treatment or control group defined by engagement level. The effectiveness of each action category was assessed by computing the heterogeneous treatment effect for each action category for each member.

Population Selection

Members enrolled in Livongo for Diabetes for a minimum of 4 months with a baseline estimated A_{1c} ≥7.5% at 30 days post enrollment were included in the study. Additional inclusion criteria were a self-reported diagnosis of type 2 diabetes at enrollment, ≥5 SMBG measures between 50 and 400 mg/dL in month one and month four of their program, and had not self-reported the use of a continuous glucose monitor (see Figure 1). Member demographics, self-reported preferences around communication and health-related interests, and level of engagement with various program features within 3 months following enrollment were used as covariates for modeling outcomes.

Figure 1. Study population funnel with inclusion and exclusion criteria. HbA$_{1c}$: hemoglobin A$_{1c}$.

Ethics Approval

The institutional review board approval was granted by Aspire IRB (#520160099), and guidelines outlined in the Declaration of Helsinki were followed.

Action Categories

Various program features were categorized and grouped to form five action categories, otherwise known as interventions per the causal inference formulations (see Textbox 1 for action category descriptions).

Textbox 1. Action category name and descriptions.

Monitoring

Number of self-monitoring blood glucose checks a member performed on the device

Coaching

Number of scheduled certified diabetes care and education specialist coaching sessions or asynchronous chat with a coach

Physical activity

Physical activity recorded using synced steps data

Nutrition

Members engagement with nutrition- and meal plan–related nudge recommendations and food log

Content

Members' engagement with educational content nudge recommendations

---

Measures

Estimated A$_{1c}$

Estimated A$_{1c}$ was calculated using the A$_{1c}$-derived average glucose model where estimated A$_{1c}$ = [mean BG over past 30 days + 46.7] / 28.7 [13]. Mean BG was calculated using SMBG values gathered through the member's device.

The intervention outcome, *Y*, is defined as the difference between estimated A$_{1c}$ in month four and month one post enrollment for members with a starting self-reported HbA$_{1c}$ ≥7.5%. Therefore, members with a more negative *Y* have a better clinical outcome.

Model Features

Treatment effect was modelled using self-reported member information and member engagement during the first 3 months post enrollment. The following variables were used as covariates in the model:

- Demographics: age, gender, BMI, race
- Self-reported medical information: self-reported HbA$_{1c}$ at enrollment, diabetes management level of self-efficacy, insulin use, on oral diabetes meds, received flu vaccine, smoking behavior
- Self-reported preferences: preferred channels of communication, interest level in becoming active and healthy
- Engagement: average days between Livongo website use; average days between Livongo mobile app use; number of

days of Livongo mobile app use; average days between SMBG checks; estimated A_{1c} at month two and three; days with SMBG hypoglycemia readings in month one, two, and three; days with SMBG hyperglycemia readings in month one, two, and three

Computing Heterogeneous Treatment Effect

The sample was defined to be composed of members with covariates, X, in treatment or control cohorts, denoted by $T=t_1$ or t_0, respectively, and intervention outcome, Y. Ideally, treatment effect, τ, would be measured for each member as:

$$\tau\,(t_0, t_1, x) = E[Y\,(t_1) - Y\,(t_0)\,|\,X=x\,]$$

However, this was not possible in a real-world data set where a member could only be in one cohort, treatment or control, at a time and not in both. Therefore, observed samples were assumed to be from a joint distribution modeled by the equations:

$$Y = g\,(T, X) \text{ and } T = f\,(X)$$

and the treatment effect was expressed as:

$$\tau\,(t_0, t_1, x) = E[g(t_1, X) - g(t_0, X)\,|\,X=x\,]$$

where $g\,(T, X)$ denoted the likelihood of outcome for a member given an intervention, and $f\,(X)$ denoted propensity of a member to be in the treatment or control cohort of the intervention.

With the assumption that all potential confounders were observed, the heterogeneous treatment effect for each member was computed using the doubly robust (DR) learning algorithm [14-17]. Treatment effect was computed by the DR algorithm using three different models. The first model performed regression on *[T, X]* to predict outcome *Y*. The second model preformed classification on *X* to predict *T*. Lastly, the two models are combined to compute the heterogenous treatment effect where the estimated outcome from the regression model is debiased by adding the inverse propensity weighted model residual. This method provided robust predictions with only one of the two predictive models needed to have a small error to obtain an unbiased treatment effect estimator [18].

Engagement Thresholds for Control and Treatment Cohort Assignment

Engagement level was used to split data into control and treatment groups. The application of the DR learning algorithm enabled intervention outcomes from the treatment and control groups to be representative of the same population because the propensity model, $f\,(X)$, incorporated any differences in population while computing the treatment effect.

Member engagement was measured for each action category during the initial 3 months post enrollment. The level of member engagement with each action category was used to assign members into the treatment or control groups through a defined threshold. If a member had higher engagement than the threshold, then the member was assigned to the treatment cohort, and the members who did not achieve the threshold were assigned to the control cohort. The treatment and control split

only included member's engagement in the action category of the program. Members in the control cohort received communication from Livongo in the form of emails and newsletters.

The engagement thresholds were defined independently for each action category. The threshold value impacted the size imbalance between the control and treatment groups, thereby affecting noise in the data set and consequently the model performance. For this reason, engagement thresholds for each action category were selected that minimize the modelling error while optimizing treatment effect.

Results

Modeling Heterogeneous Treatment Effect for Coaching Intervention

Treatment effect for the five actions categories were modelled independently. The action category of coaching is used to detail the process of selecting an engagement threshold and evaluate the heterogenous treatment effect model. Treatment effects across all action categories are then reported, followed by a proposed method to personalize action recommendations to optimize clinical outcomes.

Members who completed sufficient scheduled coaching sessions or asynchronous coaching chat sessions were assigned to the treatment cohort, with members not meeting the criteria assigned to control. This intervention engagement threshold was observed to have an impact on the DR model performance. As the threshold increases, so does the control-treatment size imbalance and noise in the data. The control-treatment size imbalance for different thresholds is shown in Figure 2.

For a member to be considered as receiving treatment, the data is split into training and validation data sets, which was split by a ratio of 65:35. The treatment effect was modelled with a forest DR learner algorithm using a gradient boosting classifier and a random forest regressor to model the likelihood of the outcome, g, and treatment propensity, f, respectively. The mean squared error (MSE) of the model was a good indicator of confidence in predicting treatment effect. A threshold value that optimizes treatment effect while having lower MSE and sufficient sample size was selected for treatment and control assignment.
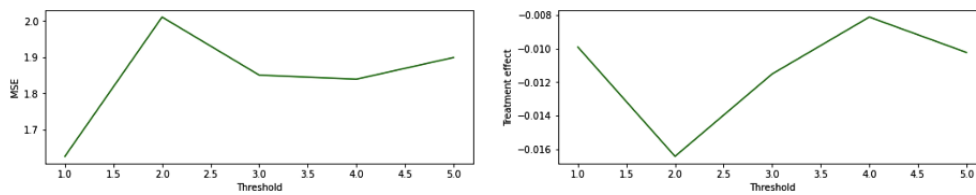
The MSE of the heterogenous treatment effect estimator model for different thresholds and computed average treatment effect are shown in Figure 3. Based on Figures 2 and 3, a threshold of at least 3 coaching sessions (scheduled or chat sessions) within the initial 3 months post enrollment was used to assign members to the treatment cohort.

Treatment effect of the coaching action category computed with the DR learner algorithm shows most members having a negative treatment effect, therefore, promoting a greater impact of coaching on estimated A_{1c} (see Figure 4). Members with positive treatment effects were those who the intervention did not improve the outcome.

Figure 2. Control and treatment cohort sample size for different minimum number of coaching sessions defined as engagement threshold.
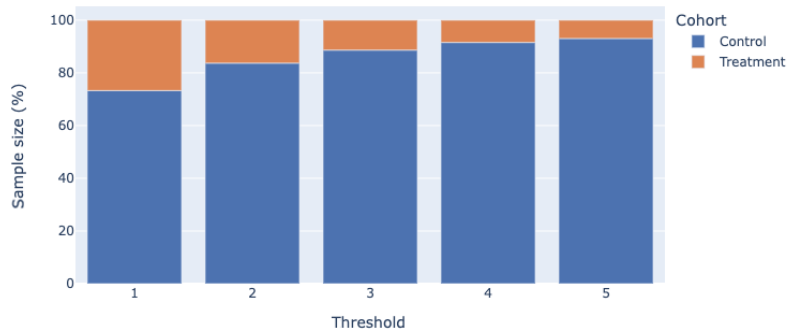
Figure 3. Top left: MSE of the doubly robust model to predict treatment effect of coaching intervention for different number of coaching sessions threshold. Top right: predicted treatment effect of coaching intervention for different thresholds. MSE: mean squared error.

Figure 4. Distribution of computed treatment effects of coaching intervention.
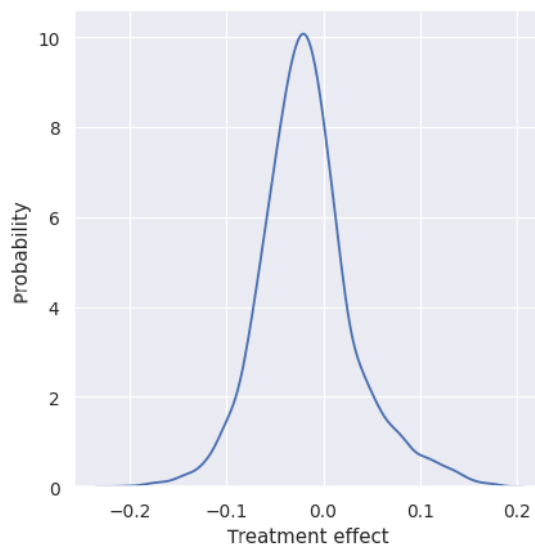
## Evaluation of Heterogenous Treatment Effect Model for Coaching Intervention

A direct evaluation of causal models cannot be made on observational data where the true treatment effect is not known due to an inability to observe the effect of being treated or not for a particular sample simultaneously. Our causal model performance was evaluated indirectly by comparing the cumulative gain of the outcome when members are ranked by model prediction when compared to random sorting [19].

Cumulative gain is cumulative uplift multiplied by sample size, where uplift is defined as the difference between average outcomes of treatment and control cohorts. A model that performs well will have large uplift values in the first quantiles and decreasing values for larger ones. By comparing the cumulative gain of members sorted by treatment effects and randomly sorted, model performance can be inferred. The higher the area under the uplift curve (AUUC) in prediction when compared to random assignment, the better the model prediction.

The cumulative gain in the outcome for the coaching action category when members are ranked by model predicted treatment effect and when randomly ordered is shown in Figure 5. The cumulative gain curves plotted are the negative of the outcome variables; therefore, the higher gain values reflect better results. The AUUC score for random assignment and assignment using the inferred treatment effect are 0.5 and 1.1, respectively.

Figure 5. Cumulative gain of coaching intervention when members are sorted by predicted treatment effect (solid blue line) and with random sorting (dashed orange line). The cumulative gain plotted is the negative of uplift from intervention outcome variable (change in A_{1c}) so that higher gain values reflect better results.
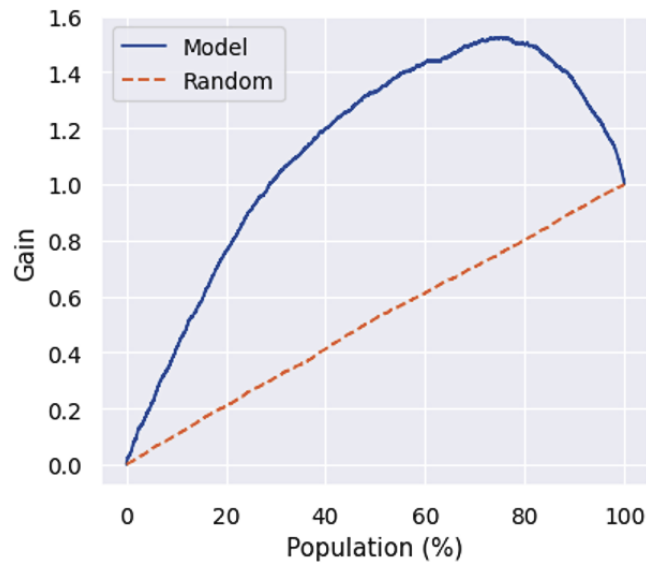
Modeling Heterogeneous Treatment Effect for Five Action Categories

The control-treatment cohort assignment for each of the five action categories for each member was inferred independently. From these independent analyses, engagement thresholds were defined for the five action categories optimizing treatment effect and sample size while lowering model MSE (see Textbox 2). The treatment and control cohorts vary across interventions depending on if they satisfied the engagement threshold condition for that intervention.

**Textbox 2.** Engagement thresholds within 90 days of enrollment for treatment.

Monitoring

≥70 days with self-monitoring blood glucose checks

Coaching

≥3 coaching sessions (scheduled coaching sessions or asynchronous chat sessions with coach)

Physical activity

≥30 days with 2000 daily steps

Nutrition

≥2 food logs or ≥50% yes responses to nutrition-related nudges

Content

≥50% yes responses to content-related nudges

Action Category Distribution and Outcomes

Figure 6 displays the total number of members within the treatment cohort of each of the five action categories and outcome *Y* related to the action category. Members with insufficient engagement in all action categories are assigned to the "other" category or control cohort.

As shown in Figure 6, the action categories had varying sample sizes. This is a result of engagement rates affected by member preferences or desired support to manage their condition. Members who engaged with coaching and physical activity had better outcomes (ie, more negative change in estimated A_{1c}).

Figure 6. Top: distribution of members in the treatment cohort within the five different intervention categories or other category as a fraction of total data set sample size. Bottom: outcome (difference in estimated A_{1c} level) for the different interventions.
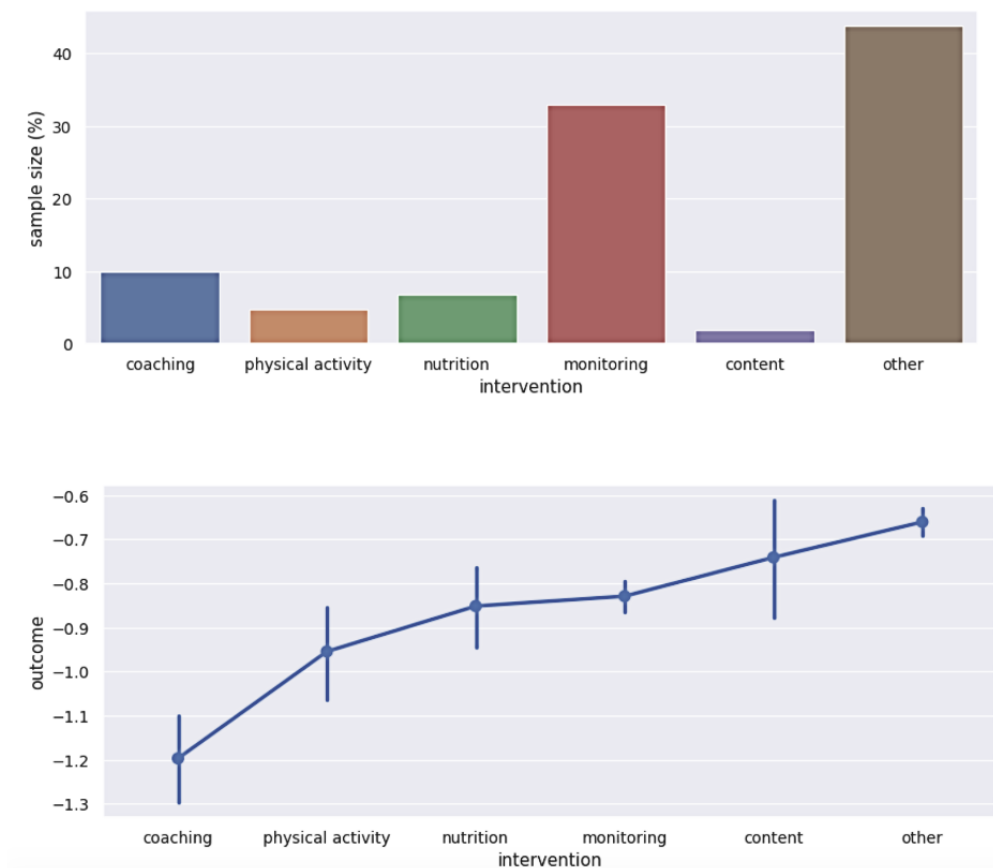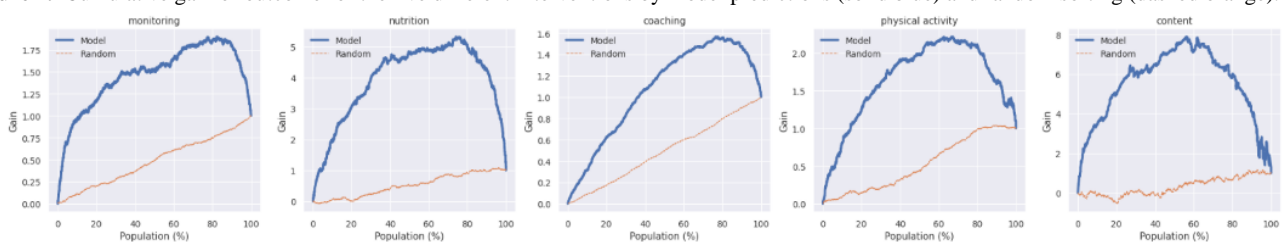

Evaluation of Heterogenous Treatment Effect Model for Five Action Categories

The model performances were evaluated using a common validation data set across interventions. The cumulative gain of the outcome when members are sorted by model predictions compared to random sorting for the five action category outcomes independently is shown in Figure 7. The area under the gain curve was larger when members were sorted by predicted treatment effect compared to random ranking, which confirmed that the model can infer causal effects of all the action categories in the data.

Figure 7. Cumulative gain of outcome for the five different interventions by model predictions (solid blue) and random sorting (dashed orange).

Action/Intervention Recommendation Based on Heterogeneous Treatment Effect

For each member, the intervention with the most negative treatment effect is the action that the model predicted would result in a larger reduction in estimated A_{1c} (ie, optimal intervention). The average change in estimated A_{1c} for members who were part of the treatment cohort in at least one action category in the validation set is shown in dark blue if the received intervention was the same as the predicted intervention and shown in light blue if the received intervention was not the same as the predicted intervention (see Figure 8). Within all five action categories, the outcome is more negative when the prediction matches the true intervention, which indicated that the model was successful in identifying optimal outcomes. Members who participated in interventions that matched their optimal predictions had an estimated A_{1c} reduction of 1.4%, while those that did not participate in their predicted optimal intervention had a reduction of only 0.57%. This 0.8% estimated A_{1c} reduction difference can be attributed to intervention personalization. Members who had a predicted optimal intervention of coaching and received coaching showed the highest change in estimated A_{1c} at 1.7%.

The distribution of predicted optimal intervention for each member with negative predicted treatment effects were coaching

(28%), SMBG checks (26%), physical activity (18%), content (16%), and nutrition (12%). Interaction with coaches and SMBG checks were observed to be the optimal intervention for 28% and 25% of the sample size, respectively, and the most recommended interventions. A balanced distribution of recommendations for optimal clinical outcomes was observed and opens an opportunity to prioritize recommendations based on a heterogeneous causal effect model.

A comparison of average treatment effect with current interventions and with recommended optimal intervention predicted by the model is shown in Figure 9. Treatment effect was larger with the recommended optimal intervention for all action categories. On average, the recommended intervention predicts a treatment effect of 0.07% compared to 0.02%, with the current intervention producing a difference of 0.05% in estimated A_{1c} attributable to personalization.

Figure 8. Change in estimated A_{1c} of members in different intervention treatment cohorts if the member received the model predicted optimal intervention (dark blue) or not (light blue).
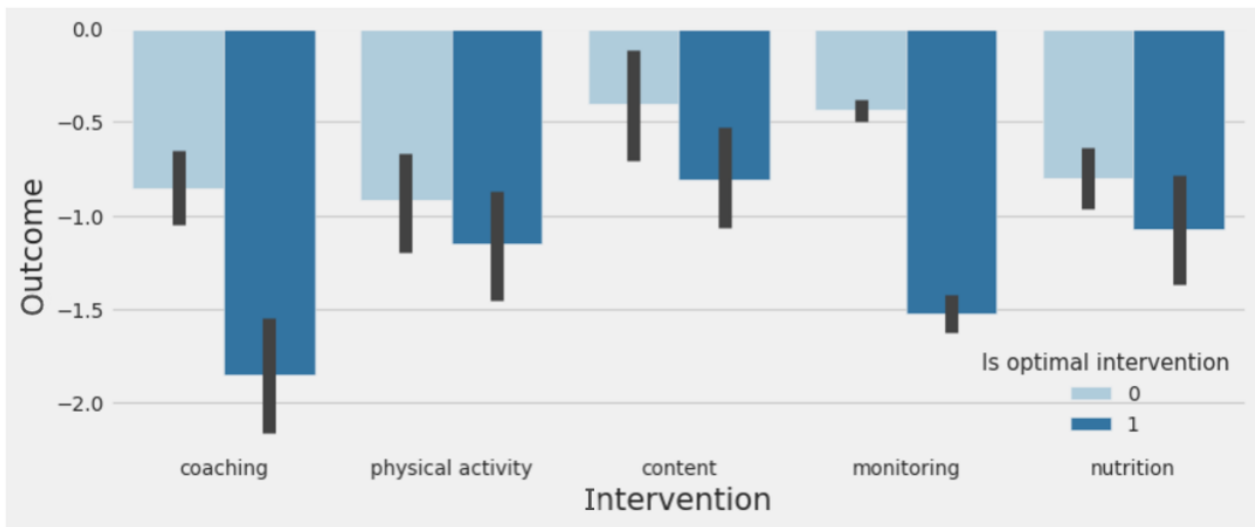
Figure 9. Computed treatment effect for members who received different interventions. The light blue bars denote the treatment effect for members who were in the treatment cohort of our data set. Dark blue bars denote treatment effect for members if they received the optimal intervention.
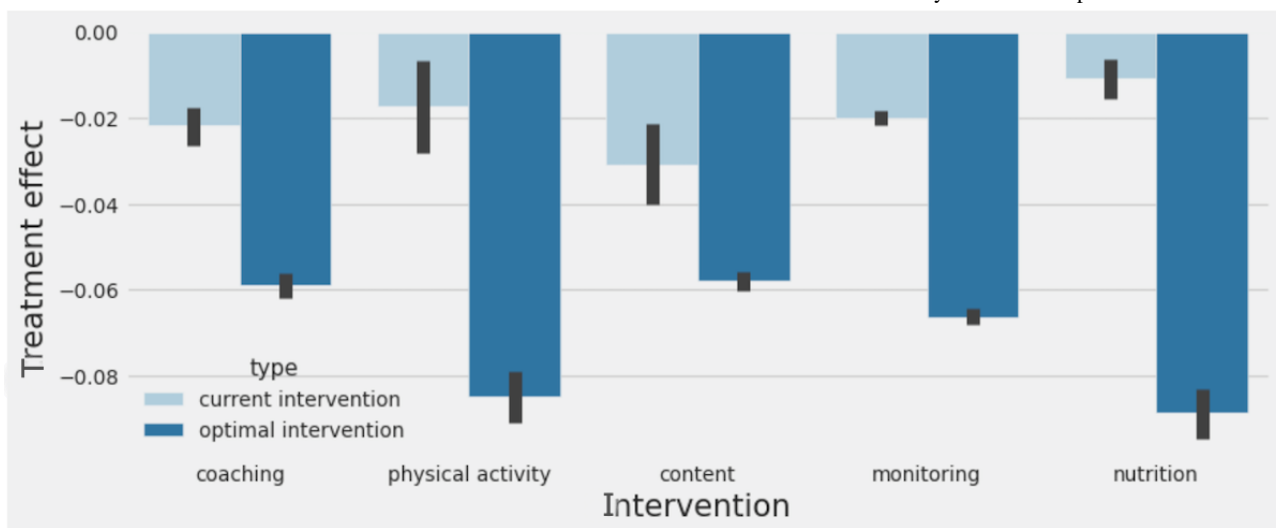
Discussion

This study highlights the feasibility of analyzing the engagement of members in an RDMP to develop a causal inference-based recommender system for predicting actions driving optimal clinical outcomes. Five action categories were identified upon member engagement level, and the causal inference model computed heterogenous treatment effect of each action per member. Model predictions were evaluated by comparing uplift gain when members were ranked by treatment effect to random

sorting, and AUUC of the model predicted gain curves were larger for all actions, validating the method to infer treatment effect. Coaching and glucose monitoring were found to be the most frequently recommended actions for members to achieve optimal clinical outcomes. On average, members who engaged within their recommended actions had a 0.8% higher reduction in estimated A_{1c} than those who did not engage within recommended actions in their first 3 months of the program, with coaching showing the largest reduction in estimated A_{1c} at 1.7% when recommended and used by members.

Machine learning has been used to study precision medicine in diabetes care and complications, variables to predict the development of diabetes, and individual characteristics related to diabetes outcomes [20-22]. However, there is a lack of evidence around practical solutions for real-world implementation, specifically around diabetes self-management behaviors, which are the foundation of successfully living with the chronic condition [23]. It has been well-established that diabetes management programs are effective in helping participants obtain glycemic control, improve HbA$_{1c}$ values, and increase self-efficacy with the support of diabetes coaches and structured SMBG [2,5,7-9]. Therefore, as real-world data has been made available around self-management behaviors through RDMPs, mobile apps, and other technologies, it is immensely valuable to use this data in machine learning techniques to provide personalized recommendations to enhance the future of diabetes care.

Our study observed better outcomes for members who engaged in their recommended actions over members who did not, across all action categories. On average, members who engaged within their recommended actions had a significant improvement estimated A$_{1c}$ than those who did not engage within recommended actions. Therefore, we propose RDMPs develop recommended actions of engagement that are more likely to lead to better outcomes based on computed heterogeneous treatment effects with the most optimal action having the most negative treatment effect. By offering personalized recommendations, members can receive a more effective experience through both digital and human coaching allowing for not only a medical cost saving for the individual through improved health outcomes but also a more cost-effective approach for the RDMP by directing the member to the most valuable program features.

Type 2 diabetes management is complex and dependent upon many factors such as nutrition, physical activity, and medication adherence, which varies widely among this population as a whole; therefore, to generate successful personalized recommendations, all variables must be gathered to match an individual's specific needs [24]. Our study presents one way that treatment effects can be computed from the causal model for recommending future interventions to drive clinical outcomes. Strong evidence around coaching, SMBG, education, nutrition, and physical activity has informed the development of diabetes management programs; however, the individual's education gaps, lifestyle, available resources, and personal priorities get lost in the wide range of program features available. This could lead to overwhelm, burnout, and even distress from not knowing where one should place their focus. By offering our method to develop a recommender system for program feature engagement, we hope that guiding members to program features that are considered most effective in supporting clinical outcomes will lessen any burdens of disease management.

This study has several strengths, including the report of real-world data, as well as insight into the demographics and program engagement of members participating in an RDMP. Members were not provided incentives to participate in the program or study beyond the Livongo for Diabetes program being provided as a benefit through their employer or health plan package. The study also had some limitations, including the retrospective analysis study design. Members in the Livongo for Diabetes program received promotional engagement outreach in the form of mobile app nudges, emails, and text messages; therefore, observational data collected for the study contained a diverse set of engagement behaviors within the program features and did not provide a clean treatment and control cohort split. Improvement in estimated A$_{1c}$ was calculated from participants' SMBG values, which has been successfully correlated with laboratory HbA$_{1c}$ values; however, it does have some limitations and is best used as a population-level tool.

Nonetheless, this study demonstrates how engagement thresholds that minimize modelling errors can be used to create control-treatment samples in observational data and compute treatment effects. The recommended action within the study is based solely on the likelihood of the member attaining a better outcome and member preferences, and propensity to engage in an action during prediction was not considered. Therefore, real-life implementation of the recommender system would have to include the likelihood of engagement and likelihood of outcome while personalizing the action recommendations. Although treatment effects were computed, it was assumed that the interventions were independent of each other and analyzed separately; however, using Bayesian inference of treatment effect would have accounted for dependencies between interventions and is recommend for a future study to further explore RDMP personalization.

Personalized action recommendations using heterogeneous treatment effects to compute the impact of member actions within an RDMP to significantly reduce estimated A$_{1c}$ can be a valuable tool in driving member behaviors toward actions that are more likely to impact clinical outcomes. Future research is recommended to implement and evaluate this model prospectively within an RDMP.

Conflicts of Interest

References

1. Risk factors for diabetes-related complications. Centers for Disease Control and Prevention. 2020 Jun 30. URL: https://www.cdc.gov/diabetes/data/statistics-report/risks-complications.html [accessed 2022-03-09]
2. American Diabetes Association. 6. Glycemic Targets: Standards of Medical Care in Diabetes-2021. Diabetes Care 2021 Jan;44(Suppl 1):S73-S84. [doi: 10.2337/dc21-S006] [Medline: 33298417]

XSL·FO

RenderX

3. American Diabetes Association. Economic costs of diabetes in the U.S. in 2017. Diabetes Care 2018 May;41(5):917-928 [FREE Full text] [doi: 10.2337/dci18-0007] [Medline: 29567642]

4. Powers MA, Bardsley J, Cypress M, Duker P, Funnell MM, Hess Fischl A, et al. Diabetes self-management education and support in type 2 diabetes: a joint position statement of the American Diabetes Association, the American Association of Diabetes Educators, and the Academy of Nutrition and Dietetics. Diabetes Care 2015 Jul;38(7):1372-1382. [doi: 10.2337/dc15-0730] [Medline: 26048904]

5. Swoboda CM, Miller CK, Wills CE. Impact of a goal setting and decision support telephone coaching intervention on diet, psychosocial, and decision outcomes among people with type 2 diabetes. Patient Educ Couns 2017 Jul;100(7):1367-1373. [doi: 10.1016/j.pec.2017.02.007] [Medline: 28215827]

6. Bollyky JB, Bravata D, Yang J, Williamson M, Schneider J. Remote lifestyle coaching plus a connected glucose meter with certified diabetes educator support improves glucose and weight loss for people with type 2 diabetes. J Diabetes Res 2018;2018:3961730. [doi: 10.1155/2018/3961730] [Medline: 29888288]

7. Gershkowitz BD, Hillert C, Crotty B. Digital coaching strategies to facilitate behavioral change in type 2 diabetes: a systematic review. J Clin Endocrinol Metab 2021 Mar 25;106(4):e1513-e1520. [doi: 10.1210/clinem/dgaa850] [Medline: 33206975]

8. Amante DJ, Harlan DM, Lemon SC, McManus DD, Olaitan OO, Pagoto SL, et al. Evaluation of a diabetes remote monitoring program facilitated by connected glucose meters for patients with poorly controlled type 2 diabetes: randomized crossover trial. JMIR Diabetes 2021 Mar 11;6(1):e25574 [FREE Full text] [doi: 10.2196/25574] [Medline: 33704077]

9. Kumar S, Moseson H, Uppal J, Juusola JL. A diabetes mobile app with in-app coaching from a certified diabetes educator reduces A1C for individuals with type 2 diabetes. Diabetes Educ 2018 Jun;44(3):226-236. [doi: 10.1177/0145721718765650] [Medline: 29575982]

10. Chrvala CA, Sherr D, Lipman RD. Diabetes self-management education for adults with type 2 diabetes mellitus: a systematic review of the effect on glycemic control. Patient Educ Couns 2016 Jun;99(6):926-943 [FREE Full text] [doi: 10.1016/j.pec.2015.11.003] [Medline: 26658704]

11. Williams DM, Parsons SN, Dunseath GJ, Stephens JW, Luzio SD, Owens DR. The impact of structured self-monitoring of blood glucose on glycaemic variability in non-insulin treated type 2 diabetes: The SMBG study, a 12-month randomised controlled trial. Diabetes Metab Syndr 2020;14(2):101-106. [doi: 10.1016/j.dsx.2020.01.006] [Medline: 31995784]

12. Gonul S, Namli T, Huisman S, Laleci Erturkmen GB, Toroslu IH, Cosar A. An expandable approach for design and personalization of digital, just-in-time adaptive interventions. J Am Med Inform Assoc 2019 Mar 01;26(3):198-210 [FREE Full text] [doi: 10.1093/jamia/ocy160] [Medline: 30590757]

13. Nathan DM, Kuenen J, Borg R, Zheng H, Schoenfeld D, Heine RJ, A1c-Derived Average Glucose Study Group. Translating the A1C assay into estimated average glucose values. Diabetes Care 2008 Aug;31(8):1473-1478 [FREE Full text] [doi: 10.2337/dc08-0545] [Medline: 18540046]

14. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. J Am Stat Assoc 2012 Feb 27;89(427):846-866. [doi: 10.1080/01621459.1994.10476818]

15. Foster DJ, Syrgkanis V. Orthogonal statistical learning. arXiv Preprint posted online on January 25, 2019.

16. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. Biometrics 2005 Dec;61(4):962-973. [doi: 10.1111/j.1541-0420.2005.00377.x] [Medline: 16401269]

17. Battocchi K, Dillon E, Hei M, Lewis G, Oka P, Oprescu M, et al. EconML: a python package for ML-based heterogensous treatment effects estimation. GitHub. 2019. URL: https://github.com/microsoft/EconML [accessed 2022-03-09]

18. Funk M, Westreich D, Wiesen C, Stürmer T, Brookhart M, Davidian M. Doubly robust estimation of causal effects. Am J Epidemiol 2011 Apr 01;173(7):761-767 [FREE Full text] [doi: 10.1093/aje/kwq439] [Medline: 21385832]

19. Gutierrez P, Gérardy JY. Causal inference and uplift modelling: a review of the literature. In: Proceedings of The 3rd International Conference on Predictive Applications and APIs. 2017 Presented at: PLMR; October 1-12, 2017; Boston, US p. 1-13 URL: https://proceedings.mlr.press/v67/gutierrez17a.html

20. Singla R, Singla A, Gupta Y, Kalra S. Artificial intelligence/machine learning in diabetes care. Indian J Endocrinol Metab 2019;23(4):495-497 [FREE Full text] [doi: 10.4103/ijem.IJEM_228_19] [Medline: 31741913]

21. Abhari S, Niakan Kalhori SR, Ebrahimi M, Hasannejadasl H, Garavand A. Artificial intelligence applications in type 2 diabetes mellitus care: focus on machine learning methods. Healthc Inform Res 2019 Oct;25(4):248-261 [FREE Full text] [doi: 10.4258/hir.2019.25.4.248] [Medline: 31777668]

22. Kaur H, Kumari V. Predictive modelling and analytics for diabetes using a machine learning approach. Appl Computing Inform 2020 Jul 28;18(1/2):90-100. [doi: 10.1016/j.aci.2018.12.004]

23. Broome DT, Hilton CB, Mehta N. Policy implications of artificial intelligence and machine learning in diabetes management. Curr Diab Rep 2020 Feb 01;20(2):5. [doi: 10.1007/s11892-020-1287-2] [Medline: 32008107]

24. Donsa K, Spat S, Beck P, Spat S, Beck P, Pieber TR, Andreas HolzingerCarsten RöckerMartina Ziefle. Towards personalization of diabetes therapy using computerized decision support and machine learning: some open problems and challenges. In: Holzinger A, Röcker C, Ziefle M, editors. Smart Health. Cham: Springer; 2015:237-260.

Abbreviations

AUUC: area under the uplift curve
BG: blood glucose
CDCES: certified diabetes care and education specialist
DR: doubly robust
HbA$_{1c}$: hemoglobin A$_{1c}$
MSE: mean squared error
RDMP: remote diabetes monitoring program
SMBG: self-monitoring of blood glucose

XSL•FO
RenderX

Original Paper

Digital Prompts to Increase Engagement With the Headspace App and for Stress Regulation Among Parents: Feasibility Study

Lisa Militello[1*], MPH, PhD, RN; Michael Sobolev[2,3*], PhD; Fabian Okeke[2], PhD; Daniel A Adler[2], BSc; Inbal Nahum-Shani[4], PhD

[1]College of Nursing, The Ohio State University, Columbus, OH, United States

[2]Cornell Tech, Cornell University, New York, NY, United States

[3]Feinstein Institute for Medical Research, Northwell Health, Great Neck, NY, United States

[4]Institute for Social Research, University of Michigan, Ann Arbor, MI, United States

[*]these authors contributed equally

Corresponding Author:
Lisa Militello, MPH, PhD, RN
College of Nursing
The Ohio State University
1585 Neil Ave
Columbus, OH, 43210
United States
Phone: 1 614 688 4248
Email: militello.14@osu.edu

Abstract

Background: Given the interrelated health of children and parents, strategies to promote stress regulation are critically important in the family context. However, the uptake of preventive mental health is limited among parents owing to competing family demands.

Objective: In this study, we aim to determine whether it is feasible and acceptable to randomize digital prompts designed to engage parents in real-time brief mindfulness activities guided by a commercially available app.

Methods: We conducted a 30-day pilot microrandomized trial among a sample of parents who used Android smartphones. Each day during a parent-specified time frame, participants had a 50% probability of receiving a prompt with a message encouraging them to engage in a mindfulness activity using a commercial app, Headspace. In the 24 hours following randomization, ecological momentary assessments and passively collected smartphone data were used to assess proximal engagement (yes or no) with the app and any mindfulness activity (with or without the app). These data were combined with baseline and exit surveys to determine feasibility and acceptability.

Results: Over 4 months, 83 interested parents were screened, 48 were eligible, 16 were enrolled, and 10 were successfully onboarded. Reasons for nonparticipation included technology barriers, privacy concerns, time constraints, or change of mind. In total, 80% (8/10) of parents who onboarded successfully completed all aspects of the intervention. While it is feasible to randomize prompt delivery, only 60% (6/10) of parents reported that the timing of prompts was helpful despite having control over the delivery window. Across the study period, we observed higher self-reported engagement with Headspace on days with prompts (31/62, 50% of days), as opposed to days without prompts (33/103, 32% of days). This pattern was consistent for most participants in this study (7/8, 87%). The time spent using the app on days with prompts (mean 566, SD 378 seconds) was descriptively higher than on days without prompts (mean 225, SD 276 seconds). App usage was highest during the first week and declined over each of the remaining 3 weeks. However, self-reported engagement in mindfulness activities without the app increased over time. Self-reported engagement with any mindfulness activity was similar on days with (40/62, 65% of days) and without (65/103, 63% of days) prompts. Participants found the Headspace app helpful (10/10, 100%) and would recommend the program to others (9/10, 90%).

Conclusions: Preliminary findings suggest that parents are receptive to using mindfulness apps to support stress management, and prompts are likely to increase engagement with the app. However, we identified several implementation challenges in the current trial, specifically a need to optimize prompt timing and frequency as a strategy to engage users in preventive digital mental health.

Introduction

Background

The stressors of parenting are normative and unavoidable. Broadly, stress is viewed in the context of life events (major or minor) that disrupt mechanisms intended to maintain one's physiology, emotion, and cognition [1]. Specifically, parenting stress is a natural experience that arises when parenting demands exceed expected and available resources [2]. Parenting stress may reflect broader contexts to include any number of major life stressors, relationships, or family circumstances [3,4]. Parenting stress can have a negative effect on relationship quality, health, mood, and overall well-being of not only the parent but also the family [3,5,6]. In both children and adults, acute and chronic stress are linked with numerous physiological and psychological disease states [5,7-9]. Thus, strategies to regulate stress are critically important among children and families [2,5-7].

However, few parents seek professional help to manage stress, and few primary care providers counsel on stress management skills [10]. Before the COVID-19 pandemic, 2019 data show that only 9.5% of adults aged ≥18 years reported receiving mental health counseling in the past year [11]. The COVID-19 pandemic took a particularly heavy toll on parents with children aged ≤18 years without much support. Evidence shows that stress management counseling is rarely offered in primary care or is the least common type of counseling provided by primary care physicians relative to diet, physical activity, or smoking cessation [10,12].

Stress management exercises (eg, mindfulness, cognitive reframing, and behavioral modification) are empirically supported strategies to promote self-regulatory skills, stress reduction, and resilience [13-15]. Evidence suggests that mindfulness interventions have the potential to reduce parenting stress and improve psychological functioning in youth [5]. Specifically, mindful parenting (ie, moment-to-moment awareness of the parent-child relationship) may reduce parental stress and promote family well-being [6,16]. However, it is unclear how formal (ie, purposeful or dedicated time) or informal (ie, weaving mindfulness into existing routines such as dishwashing) mindfulness is required to yield positive outcomes [17,18]. Mindfulness intervention dosage has wide variability reported anywhere from 9 to 27 hours [5], 45 minutes to 2 hours delivered over 3 to 12 sessions [19], and 4.5 to 24 hours [20]. Collective findings from systematic reviews and meta-analyses suggest small but beneficial effects of mindfulness interventions [5,19,20].

However, American families are often busy citing time and logistical barriers (costs and transportation) as reasons for attrition or nonparticipation in health-promoting activities. Both before and during the COVID-19 pandemic, families struggled to balance work and life, often feeling tired, rushed, and short on quality time with their children, friends, and hobbies [21]. In 2019, around 51% of mothers and 82% of fathers reported working full-time. In 2015, 26% of children aged <18 years lived with a single parent, and 53% to 73% of parents reported that their child participated in an extracurricular activity in the previous 12 months [21].

Brief mindfulness interventions may be a suitable alternative to more traditional mindfulness programs for populations with limited time or available resources [22]. In contrast to traditional 8- to 10-week mindfulness interventions, brief mindfulness interventions range from 3- to 5-minute guided exercises to 2-week programs [22]. Brief mindfulness interventions have also become increasingly popular with the rise of commercially available mindfulness meditation apps and other technologies [22-25]. Both commercial and research-developed mobile apps that help participants learn, practice, and monitor mindfulness activities [17,18,25,26] may increase opportunities to learn and practice mindfulness or meditation (hereinafter collectively referred to as mindfulness) across diverse audiences compared with traditional face-to-face programs [27], and successful coping strategies may be deployed in real time in real-world conditions to mitigate the deleterious effects of parenting stress [2,5,28].

A critical challenge in digital health, the law of attrition, occurs when digital health study participants drop out before completion or stop using the app [29,30]. Interventions intended to revolutionize stress regulation, including digitally supported health interventions, often fail because they are not used in real-world settings [31-35]. The main users of mental health apps are predominantly younger individuals, of high socioeconomic status, have overall positive health, and routinely engage in preventive health practices [36-39]. Families, particularly of low income, have poor adoption rates of abundant mobile health (mHealth) solutions despite high smartphone ownership and potential for scalability [32,33,38]. Poor engagement with digital health interventions may be mitigated through real-world strategies [31,40,41]. Many mHealth systems lack sufficient effort devoted to their design, development, and evaluation across diverse populations [15,31,34]. A solution-focused approach prioritizes the development of a solution to a practical problem to produce a sustainable solution [40,42].

A critical first step in supporting parental stress regulation (distal outcome) was to identify whether and under what conditions prompting parents to engage in mindfulness is beneficial. Digital prompts (eg, push notifications) are frequently used to promote engagement. Digital prompts are intended to nudge users in a particular direction without limiting the freedom of choice [43-45]. However, an overflow of notifications can be burdensome, leading users to ignore push notifications, increase user inattention, and/or exacerbate disengagement from apps [30,46]. Digital prompts are subject to a myriad of factors such as timing, frequency, sender, content, message framing, mode

XSL•FO

RenderX

of delivery, and theoretical underpinning [43-45]. Previous research on the use of digital health prompts to increase engagement with interventions among parent populations is limited [47,48]. One study conducted with parents found that the timing of prompts to support healthy lifestyle behaviors may be beneficial if delivery coincides with parents' perceived need for support (eg, prompt to practice guided imagery sent when the parent was at home versus while driving) [47]. Parent preference for prompt timing (ie, self-selected time frame of when to receive a prompt) peaked during late afternoon and evening hours when the school or workday ended and the family transitioned to dinnertime and evening activities [47]. With regard to prompt frequency, survey data show that parents desire few notifications (ie, 2 times per week) [48], while intervention data suggest that parents favor frequent messages (eg, daily) [47,48]. As mobile apps evolve, opportunities for meaningful intervention within life patterns may be possible through various mechanisms that sense and capture streams of personal data, giving insight into contextual factors [43,44,49-51].

Just-in-time adaptive interventions (JITAIs) have potential for shaping health behavior, using various data streams to deliver prompts at the *right time* while minimizing user burden and habituation [52]. JITAIs rely on explicit decision rules for when to prompt users with specific intervention components [52]. However, it is uncertain when and under what conditions, interventions should be delivered to engage parents in real-time stress-regulating activities. A microrandomized trial (MRT) is an experimental design that can supplement the use of theory to guide JITAI development. An MRT allows researchers to study the proximal effects of a specific intervention component, change over time, and contextual factors that may moderate time-varying effects [52].

Therefore, before designing a comprehensive JITAI for parenting stress, we conducted a pilot MRT to explore whether it was possible to leverage digital prompts to support real-time parent engagement with stress-regulating activities guided by a commercially available mindfulness app, Headspace. We developed a novel system that, from the front end (parent perspective), leveraged digital prompts containing messages that encouraged brief (<10 minutes) mindfulness activities and were capable of launching Headspace. Parents were able to change the timing of prompt delivery to suit their individual needs through a secondary research-developed app. From the back end (researcher perspective), the system would randomize to send or not send a digital prompt during the parent-specified window. Headspace was used to teach and support mindfulness activities. While adaptability and scalability were facilitated using a commercially available app, proprietary back-end data for Headspace were not accessible by our research team (ie, what content parents accessed on Headspace). Therefore, our system was designed to passively capture relevant smartphone data (eg, when the Headspace app was open; how long the app was open; or smartphone paradata such as silent, ring, or vibrate mode). Finally, our system pushed a daily ecological momentary assessment (EMA). The EMA was used to differentiate engagement with the app versus engagement with mindfulness and to observe for potential benefits on parent mood.

Objective

In this study, we aim to conduct a pilot MRT to determine if it was feasible and acceptable to randomize digital prompts designed to engage parents in real-time brief mindfulness activities guided by a commercially available app (Headspace).

Methods

Trial Design

We conducted a 30-day feasibility and acceptability study whereby participants were microrandomized [53] once daily with equal probability to either receive or not receive a prompt recommending engagement in a commercially available mindfulness app (Figure 1). Microrandomizations took place within a time window that each participant prespecified as convenient to receive the prompt. Specifically, during the onboarding session, participants were asked to specify a 3-hour window during the day to receive a prompt encouraging them to practice mindfulness. On the basis of our prior work [47], a 3-hour window was thought to be broad enough but not so restrictive to capture windows of family routines (eg, morning before work or school, work or school lunch or breaks, after work or school dinner, evening activities, or bedtime routines). The time frame can be updated at any point by the user. If a prompt was sent, the participant could either tap the prompt and launch a mindfulness app or dismiss the prompt. Prompt messages were neutral in tone and picked at random from a library developed for the study (eg, "Take your mind to your calm place"; "Use mindfulness to develop your mental toughness"; "10 minutes for yourself can make a huge difference"; and "Give yourself time to recharge"). In the 24 hours following randomization, proximal outcomes of engagement in mindfulness exercises and affect were self-reported via daily EMAs and passively assessed via the smartphone.

Figure 1. Trial design.
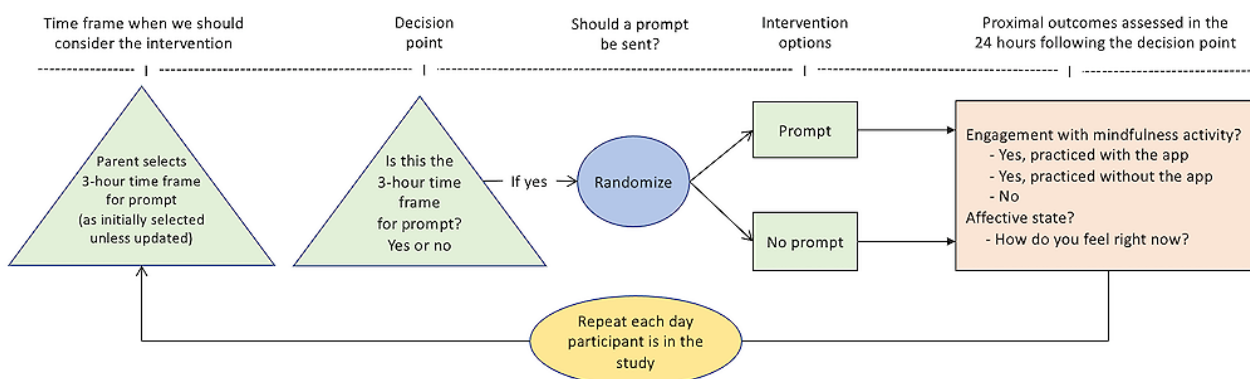
Procedure and Ethics Approval

From November 2018 to February 2019, participants were recruited from web-based or social media announcements, word of mouth, and targeted community and workplace recruitment (eg, email notifications via listservs). Announcements included information regarding the study, contact information, and a link to a 3-item eligibility screening survey. Owing to technical aspects of the platform, participant inclusion criteria were as follows: (1) self-identified parent with child or children (aged up to 18 years) at home and (2) an Android smartphone user. The exclusion criterion was non–English-speaking participants. Upon meeting the inclusion criteria, participants provided written informed consent and were enrolled in the study. The setting was the real-world, everyday lives of the participants. Given the formative nature of the proposed work, a convenience sample was used. In appreciation of participant time and feedback, parents were compensated commensurate with participation. Parents were provided compensation for baseline survey completion (US $5), onboarding (US $5), EMA (up to US $15), exit survey (US $5), and a 1-month Headspace Plus paid subscription (valued at approximately US $13). The lead

institution was the Ohio State Behavioral and Social Sciences Institutional Review Board (IRB) with ethical approval from the Ohio State University Office of Responsible Research Practices (IRB #2017B0550).

After obtaining informed consent and completing baseline surveys, participants were provided with onboarding instructions. As part of feasibility testing and formative evaluation, participants were asked to install 3 mobile apps necessary to conduct the research (Figure 2). Two research-developed apps (Beehive and App Logger) were made available to participants through a provided link, while one app (Headspace) was commercially available for download through the Google Play Store. It is worth noting that the research team submitted a request for collaboration with Headspace but owing to timing and resource limitations, the Headspace Health Partnership team was unable to accept our request. However, the principal investigator (PI) communicated with a member of the Headspace team to discuss logistical questions related to onboarding participants. Upon study completion, participants were asked to complete exit surveys and uninstall the research-affiliated apps. Given the commercial availability of Headspace, this app could be continued at the user's discretion.

Figure 2. Research apps Beehive, HeadspaceTM, and App Logger.
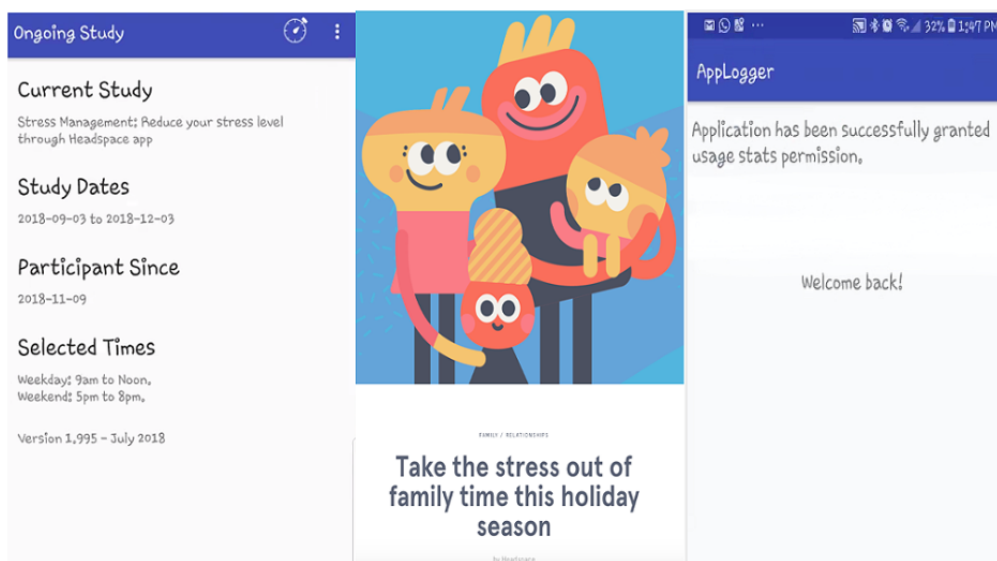
Mobile Apps

Beehive

Beehive is a research-developed app that participants installed to guide the delivery of study prompts [54] and the EMA questions. Participants used Beehive to indicate daily wake and bedtime schedules and allowed participants to select the preferred timing for the digital prompts. Specifically, participants were instructed to use the Beehive app to select a 3-hour window to potentially receive intervention prompts encouraging them to practice mindfulness. Parents were informed that they could update the 3-hour window through the Beehive app at any point during the study. The Beehive app also delivered the EMA question at the end of the day and at least one hour before the specified bedtime of each participant.

Headspace App

Prior research has leveraged commercially available tools to support stress regulation and optimize digital health interventions [26,44,55]. We selected Headspace, a commercially available mobile app that includes a wide collection of mindfulness exercises that vary in terms of length and topic [56]. In a review of mindfulness-based iPhone apps and using the Mobile Application Rating Scale [55] to determine app quality, Mani et al [24] found Headspace to have the highest average score. Participants were provided with a code that gave them access to the full Headspace library and were asked to use the app over the course of 1 month. Headspace includes push notifications that participants were asked to disable during the study.

App Logger

App Logger is a research-developed app used to facilitate data collection. It is an unobtrusive mobile app for Android devices that passively records and timestamps smartphone paradata [54]. App Logger was used in prior studies to objectively measure app usage but not content [54,57]. It was used to passively capture smartphone paradata, such as when an app was launched, duration of use (removed from the foreground), and smartphone mode (eg, locked or unlocked). In lieu of a partnership with Headspace, App Logger was used to passively collect relevant smartphone paradata and contribute to a holistic view of engagement.

Measures

Feasibility and Acceptability

Feasibility and acceptability were assessed via participant enrollment and retention rates, satisfaction and acceptability ratings (a benchmark of ≥90%), estimates of use, self-reported engagement with mindfulness exercises (via EMA) and objective engagement with the app (ie, passively collected app usage, patterns, and trends over time), reactions to the intervention, factors affecting implementation ease or difficulty, and the ability of participants to carry out study activities [58]. A research log was used to document the proportion of eligible parents, relative to those who enrolled and subsequently those who enrolled compared with actual attendance. These data were used to inform whether we were able to recruit our target population. We determined whether randomization was feasible

by monitoring software performance and was acceptable by monitoring for aberrant data. Retention rates and reasons for nonparticipation were obtained when possible. A database of prompts sent, date, and timestamps was maintained. Additional smartphone paradata, information regarding smartphone use, were assessed using the App Logger app. App Logger passively records when a prompt is delivered, time from delivery to launching the Headspace app, and duration the Headspace app is open (not content). In addition, App Logger can detect the smartphone mode and setting (eg, on or off or silent, vibrate, or ring mode). We have reported on all study outcomes to determine the extent and likelihood that the intervention was implemented as planned.

Baseline Survey

The baseline survey collected basic demographic information and previous experiences with mindfulness. In addition, we assessed for the personality factor, neuroticism, using the neuroticism subscale of the Big Five Personality Inventory [59,60]. Neuroticism is an indicator of emotional stability, with higher scores suggesting a higher likelihood of vulnerability to stress, anxiety, and nonspecific mental distress [61-63]. The neuroticism subscale consists of 8 items rated on a 5-point scale where 1=disagree, 3=neutral, and 5=agree. To assess depression, anxiety, and stress, we used the Depression Anxiety Stress Scale 21-item (DASS-21) [64]. The DASS-21 has been successfully used in a variety of clinical and nonclinical settings, including parent populations [65-67]. Participants answered each item based on how it applied to them in the past week. Responses ranged from 0=did not apply to me at all to 3=applied to me most of the time. The DASS-21 cutoff values indicate levels such as normal, mild, moderate, severe, and extremely severe within each state; that is, depression, anxiety, or stress.

Ecological Momentary Assessments

Nightly EMAs were scheduled for delivery within 1 hour of participant-indicated bedtimes. We used the Photographic Affect Meter (PAM [68]) to measure the daily affective state. The PAM typically takes users less than a minute to complete, asking users to choose from a grid of 16 photographs arranged in a 4 × 4 grid that most represents their current state. Each image was mapped to the established valence and arousal states. Participants were asked to "touch the photo that best captures how you feel right now." The output of this selection is a positive or negative affect value, which has been validated in multiple trials and shown to be an effective alternative to longer-form surveys [68,69].

Self-reported engagement in mindfulness exercises was measured by asking participants whether they had performed a mindfulness activity. Response options were "Yes, I used the app," "Yes, I practiced on my own," or "No." These options were provided to ascertain how participants engaged in mindfulness exercises, with versus without the app.

End-of-Study Survey

A 14-item survey consisting of both open- and close-ended questions was used to elicit participants' overall thoughts about the study, prompt-specific feedback, and Headspace-specific feedback (ie, prompt delivery, prompt content, and Headspace

app). Survey questions were about general perceived ease of use, usefulness, and satisfaction based on input from prior work with parent populations [47,48], usability [70], and expert recommendations for feasibility and acceptability studies [58,71].

Examples of yes or no questions included the following: (1) Was it easy to get set up and started? (2) Did you dismiss any notifications? (3) Did you find the overall program useful in managing your stress? (4) Would you recommend the program to other parents? Prompt-specific questions included questions such as the following: (1) Was the number of messages sent (frequency) too much, too little, or just right? (2) Was the time when the prompt was delivered helpful, not helpful, or just right? (3) Did you like the wording of the messages: yes or no? (4) Did your feelings toward the notifications change the longer you were in the study? (5) Was there anything that we could have done better? Headspace-specific questions included the following: (1) Did you find the Headspace activities helpful in managing your stress? (2) Did you like the graphics and characters used in Headspace? (3) Can you list one Headspace activity that you liked most and least? (4) Do you think it would be helpful for your kids to use Headspace to learn mindfulness?

Data Evaluation

Feasibility and acceptability were assessed via participant enrollment and retention rates, satisfaction, factors affecting implementation ease and difficulty, patterns of use, and ability of participants to carry out study activities [58]. The proximal outcome, engagement with a mindfulness activity, was assessed in the 24 hours following randomization (when the system would randomize to send or not send a prompt). Engagement in a mindfulness activity was operationalized to disentangle engagement with the app (eg, simply opening the Headspace app) versus engagement with a mindfulness activity (eg, deep breathing with the app or deep breathing on my own without the app). Engagement in any mindfulness activity was assessed along with the affective state (transient emotions) from daily EMAs.

Quantitative data were analyzed descriptively. Benchmarks for recruitment and enrollment efforts were set at ≥50 participants screened and 20 participants enrolled (most studies had a sample of <50 participants; nearly one-third—ie, 6/2, 30%—of studies had a sample of <20 participants) based on a systematic review of similar research [20] and given the pilot nature of the proposed work. Additional indicators of recruitment success

were if ≥24% of the enrolled participants represented racial or ethnic minorities [20]. Owing to limited supporting evidence, we also reported on frequencies of participation from low-income families and fathers (compared with mothers) to help fill the gaps in knowledge. We anticipated 100% success in onboarding and aimed for <20% attrition during the 30-day intervention implementation. We report on response rates to EMAs as we have done in our prior work, with a benchmark for success calculated at 60% or higher (response rate 49% for 35 days [47] and 61% for 30 days [72]).

Qualitative data from free-response questions was narratively summarized by 2 coders representing 2 different disciplines. The study PI (LM) served as the lead coder based on the population and subject matter expertise. The second coder (MS) provided topical expertise on the subject and helped to verify the findings against actual data. Any areas of concern were discussed. The process began with an initial reading of the data before coding. Multiple readings of free-response data were performed to identify key words or phrases. Passages coded similarly were grouped into themes to provide a more comprehensive view of the data.

Results

Sample

Over the course of 4 months, we screened 83 parents who expressed interest in the research. Approximately 42% (35/83) of the parents were excluded before conducting the research, mainly because they were not Android users. Of the 48 eligible patients, 16 (33%) parents were enrolled. Figure 3, the CONSORT diagram, depicts enrollment, participation, and analysis. Most of our sample identified as White and female and were aged between 35 and 44 years. All participants used Android devices as their primary mobile smartphone, with Samsung being the predominantly reported brand (6/10, 60%). At baseline, the sample average score for neuroticism was moderate (approximately 26 out of 40), suggesting a sample not particularly calm nor prone to emotional instability. Similarly, assessments of depression, anxiety, and stress were largely within normal ranges. Although most of the participants reported practicing stress management (13/16, 81%), fewer reported using mindfulness (6/16, 38%) or the Headspace app (1/16, 6%). Table S1 in Multimedia Appendix 1 shows the sample characteristics.

Figure 3. CONSORT (Consolidated Standards of Reporting Trials) diagram.
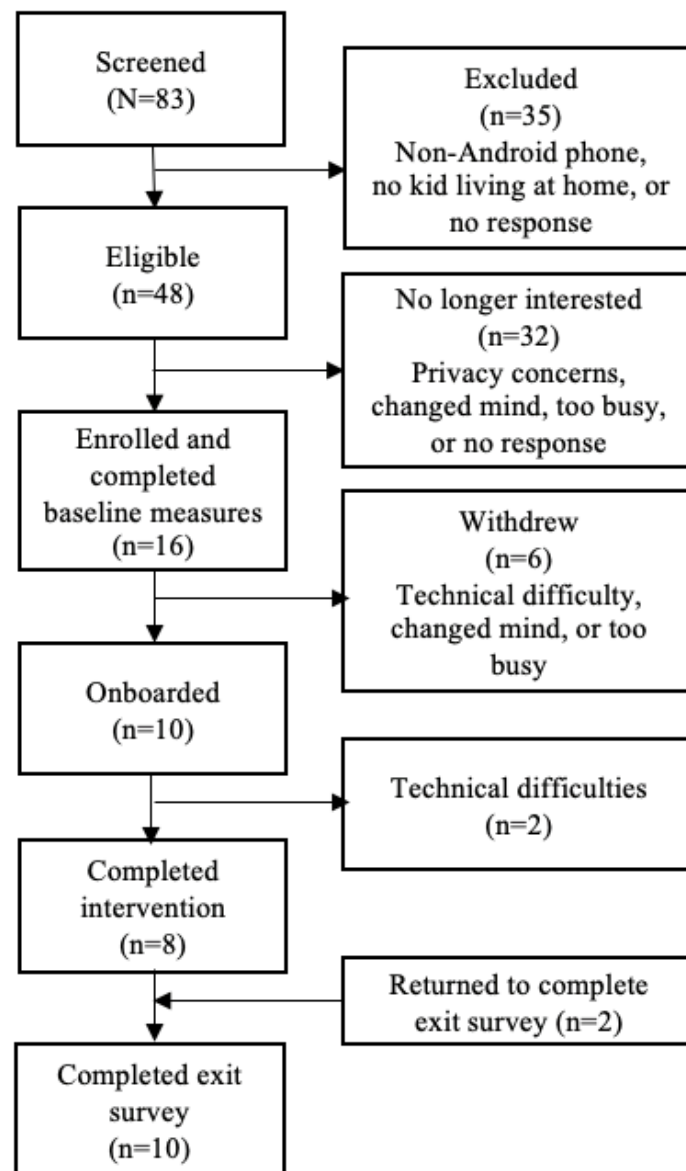
Feasibility

Recruitment and enrollment were conducted on a rolling basis, but intervention implementation was grouped into waves to prevent overlap with the winter holiday season and to meet logistical needs associated with deploying Headspace bulk subscriptions. As such, most participants were screened and enrolled during the busy winter holiday season (November-December: 12/16, 75%), while efforts slowed after the New Year (January-February: 4/16, 25%). In addition, 12% (2/16) were fathers, 12% (2/16) reported receiving medical assistance, and 6% (1/16) represented a racial or ethnic minority. We were able to meet our recruitment goal but were shy of our enrollment goal and unable to meet our goal to have ≥24% of racial or ethnic minority participation. Enrollment was further aggravated with reductions in sample size occurring during the onboarding process.

Collectively, 50% (8/16) of parents reported technical difficulties during the onboarding process, but 25% (2/8) were

able to overcome these barriers after speaking with a member of the research team. In wave 1, 33% (4/12) of participants were lost during onboarding, and in wave 2, a total of 50% (2/4) of participants were lost during onboarding. Logistical challenges during onboarding can be broadly categorized into two categories: (1) integrating a commercial app into research and (2) using research-developed apps in real-world settings. First, it was necessary for the PI to collaborate with the Headspace team to navigate the logistics of how to purchase and transfer separate Headspace subscriptions to each participant. At the time, Headspace allowed for bulk subscription purchases, akin to a corporate account, but required at least 30 subscriptions to be purchased at once. Furthermore, bulk subscriptions had to be activated on the same day, which was counter to the rolling recruitment research protocol. This was solved in 2 ways. We arranged a small cohort of participants to start the study on the same day using a promo code associated with a bulk subscription. In addition, we provided individual Headspace gift subscriptions to participants via rolling recruitment. Both

strategies allowed parents to download the Headspace app from the Google Play Store and enter a promo code to unlock the premium version of Headspace. Most parents were familiar with the Google Play Store and located the Headspace app for download. In total, 3 of 16 (19%) parents reported that the Headspace promo code did not work when they were onboarding. During these scenarios, the PI helped troubleshoot to resolve the issue. The second challenge was the implementation of the 2 research apps, both available via GitHub. Participants were provided with a direct link to download the research apps. Although the research apps allowed the research team to have greater control, some smartphones triggered a warning during installation. Such warnings may be standard on some Android devices when an app is downloaded from a location other than the Google Play Store. However, parents reported being unfamiliar with downloading apps from outside the Google Play Store:

> I can't download apps from an outside source that isn't Google Play. Unfortunately it looks like I won't be able to participate in the study. [P5]

Similarly, parents voiced concerns regarding privacy and smartphone access, although they were informed that the content was not being monitored:

> I'm fine with it tracking Headspace usage, but I'm a bit concerned if it's monitoring all my app, location, calls or/and keyboard usage. [P7]

> It's like asking someone to put a chair in your living room, even more so it's like asking someone to move in, especially when they are requesting access on your phone. [P4]

Ultimately, 63% (10/16) of the enrolled parents onboarded to the 30-day intervention phase of the study. However, incoming data streams from App Logger were interrupted for 2 of 10 (20%) participants. We were unable to ascertain with certainty whether these interruptions were due to system issues or whether these participants chose to alter or revoke App Logger permissions. Thus, 8 of 10 (80%) parents fully completed the intervention phase and provided the following insights. An average of 10 (SD 5.44) prompts per participant was delivered over the 30 days relative to intended on-average 15 prompts (ie, 50% chance of receiving a prompt × 30 days in the study). Prompt messaging may be found in Multimedia Appendix 2. Across the 8 parents, smartphones were set to the normal ring mode 46% of the time when a prompt was sent (44 observed instances where the smartphone was set to the normal ring mode/95 prompts sent). The average time from prompt delivery to action (eg, either launch Headspace or dismiss notifications) was 185 (SD 303) minutes. A reliable reference point for nonprompt to action was not available; therefore, a comparison analysis was not possible. Figure 4 shows the hour of day (from 0 to 24 hours, after midnight) when a participant opened the Headspace app. The x-axis separates participants, shows the mean as the point, and shows the 95% CI around the mean as error bars. From the 95% CIs, we see visually that the variability in the time of day of opening Headspace differed, with some individuals using Headspace at a more consistent time of day than others.

Proportionally, most prompts were dismissed if the smartphone was in the silent mode (Table 1). The response rate to EMA surveys was 72.4% (ie, 165 EMA responses/228 EMAs pushed).

Figure 4. Distribution of hour of day (from 0 to 24, after midnight) when a participant opened the Headspace app.
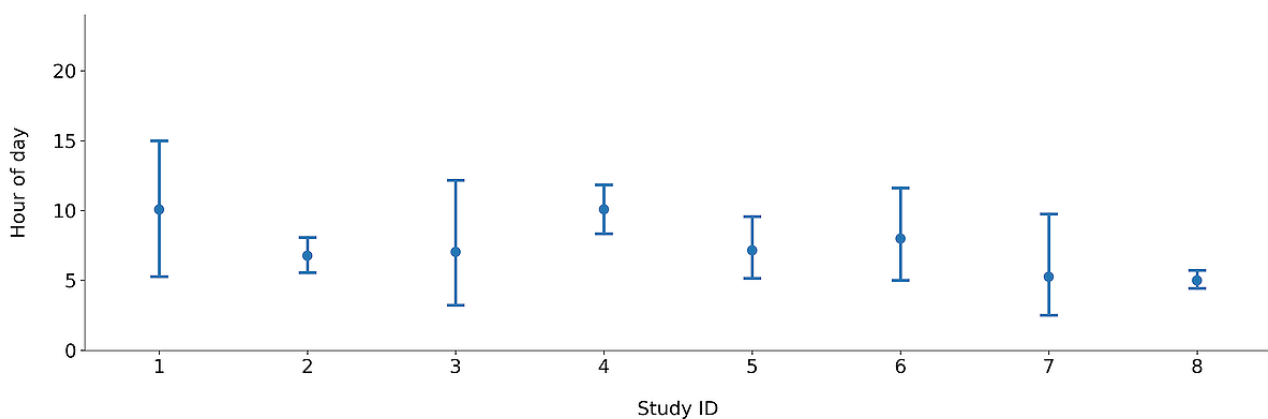
Table 1. Smartphone ringer mode when the prompt was sent.

Ringer mode	Value (n=82), n (%)	Was the prompt dismissed?[a]	
		Yes (by mode), n (%)	No (by mode), n (%)
Normal	39 (48)	22 (56)	17 (44)
Silent	12 (15)	9 (75)	3 (25)
Vibrate	31 (38)	18 (58)	13 (42)

[a]The "Value" column was used as the denominator.

Engagement

Engagement With the Headspace App

Over 30 days and across the 8 participants, we counted 298 engagements with the Headspace app from passively collected smartphone data. We noted higher self-reported engagement with the app on days when a prompt was sent (31/62, 50% of days) compared with days without a prompt (33/103, 32% of days). Across the study, most participants (7/8, 87%) in the study self-reported higher engagement with the app on days

with prompts. Engagement with the app was also found to be longer on days with the prompt (mean 566, SD 378 seconds) than on days without a prompt (mean 225, SD 276 seconds). Across the 8 participants, we observed that most engagements with the Headspace app occurred during the first week of the study but subsequently tapered down by week 4. Figure 5 highlights the total duration of engagement with the app by week in the study across the entire sample. The high error bars in weeks 1 and 2 suggest a larger variation in individual use during the first 2 weeks compared with the last 2 weeks.

Figure 5. Total duration of participant engagement with app by week in study across all participants (n=8). Points represent the mean value, and error bars around each point represent a 95% CI around the mean.
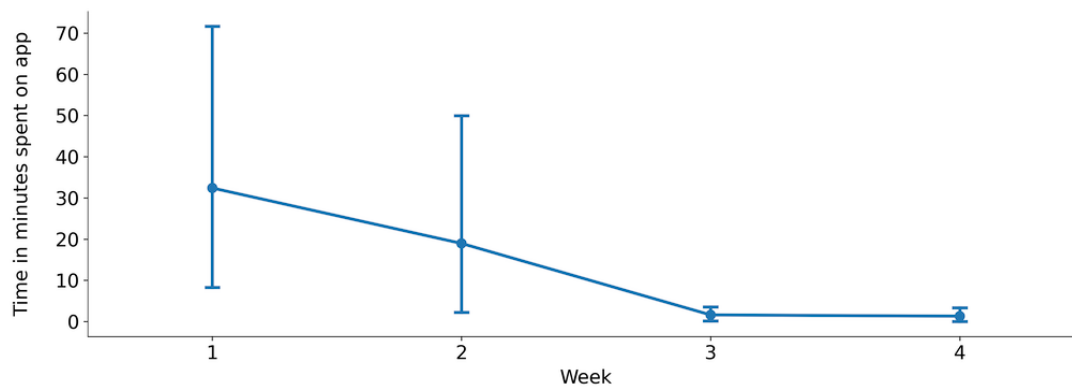
The median (IQR) number of engagements with the app per user was 19 (10-55). The median (IQR) duration of each log-in was 45 (15-110) seconds. We discovered 25 discrepancies between reported (yes or no) and observed (yes or no) app usage, where parent-reported data from EMAs indicated app usage, yet we found no recorded Headspace app use on passively recorded data. On exit surveys, 50% (5/10) of parents reported using the app for an average of 5 to 10 minutes per day, and 40% (4/10) reported using the app 1 to 3 days per week.

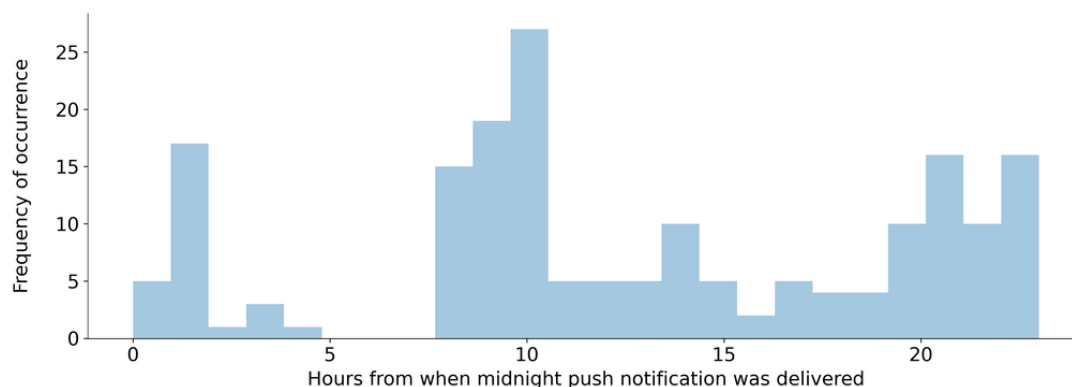#### Self-reported Engagement With Mindfulness Exercises (Based on Daily EMAs)

On the basis of when prompts were sent, we observed a multimodal distribution for the time of day when the prompt

was sent, most often during the morning hours, peaking between 9 AM and 10 AM (Figure 6). Proportionally, parent-reported engagement in a mindfulness activity (Table 2) was descriptively similar on days when a prompt was not sent (65/103, 63%) compared with that on days when a prompt was sent (40/62, 65%).

Collectively, parents reported using the Headspace app for mindfulness activity (64/105, 61%) more than engaging in mindfulness activities without the app (41/105, 39%; Table 2).

We observed a change over time. During weeks 1 and 2, parents reported using the Headspace app to support mindfulness activities. However, in weeks 3 and 4, mindfulness activities without the app increased (Figure 7).

Figure 6. Hours from midnight prompts were sent based on the parent-selected time frame. Each bar represents a single hour of the day, and the height of the bars represents the total number of push notifications sent at that specific hour across all participants in the study.

Table 2. Self-reported engagement with any mindfulness activity across all users (n=8).

Parameters	No prompt	Prompt	Total
Ecological momentary assessments collected, n (%)	103 (62.4)	62 (37.6)	165 (100)
Engagements with any mindfulness behavior (with or without the app) indicated by parent report, n (%)	65 (63.1)	40 (65)	105 (100)
Engagements with mindfulness activity using the Headspace app indicated by parent report, n (%)	33 (32)	31 (50)	64 (100)

Figure 7. Self-reported engagement with a mindfulness activity by week in study (n=8). Points represent the average number of ecological momentary assessment (EMA) responses within a week across participants, and error bars represent a 95% CI around the mean value. Each trajectory represents a specific type of EMA response (eg, "Mindfulness without Headspace") as indicated in the legend.
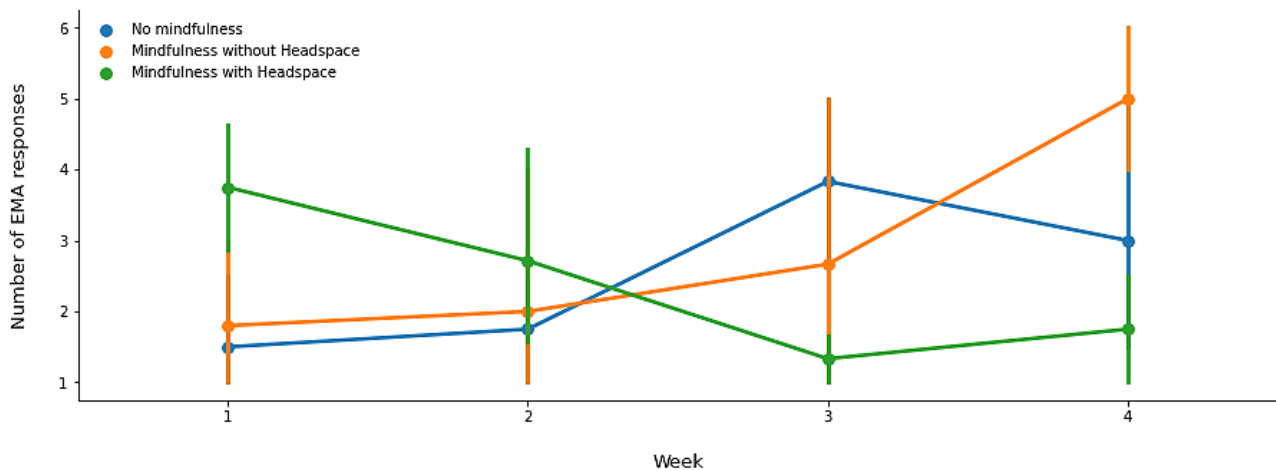
Emotional State or Affect

Across the 8 participants, the top three emotional states captured from the PAM responses were tired, satisfied, and sleepy. We did not observe any patterns or trends in emotional state over the course of the study, nor did we observe any trends between affect and engagement.

Acceptability

Acceptability and parent satisfaction were determined from free-response text on baseline and exit surveys, as well as from communication with parents (eg, during onboarding). Although 2 participants were affected by technology errors impacting EMA collection, they were included in exit surveys (n=10) as their experience could provide additional insight into outcomes. Overall, only 70% (7/10) found the program useful in managing stress, which is below our satisfaction benchmark of 90%. However, 90% (9/10) found the research helpful and would also recommend the program and Headspace app to a friend.

Specifically, most participants liked the wording of the message prompt (9/10, 90%) yet reported dismissing the prompt at some point over the course of the study (8/10, 80%). When specifically asked about the exit survey, 50% (5/10) indicated that the messages could have been more caring (eg, "We care about your health, take a min to manage your stress") or humorous ("funnier"), but not authoritative ("do this now" type of message; 1/10, 10%). The majority (7/10, 70%) did not believe a visual within the prompt (eg, emoji or meme) would be helpful. Participants most often viewed prompts as reminders (9/10, 90%), while some participants (4/10, 40%) reported that

prompts became "more annoying the longer I was in the study." Just over half of the participants reported that both the timing and frequency of prompts were helpful (6/10, 60%). However, responses to open-ended questions highlighted the competing demands for parent attention in the moment:

> I got a notification while I was out running errands. [P9]

> It felt like one more to-do on a long list of to-dos. I occasionally wanted to do mindfulness out and about while waiting for something, but I couldn't download them for offline use. [P3]

> Messages that were trying to be nice felt very tone deaf when circumstances conflicted with them, ie "the stressful part of your day is done" popping up on my phone as my baby is screaming. [P4]

> I had tons of other notifications and I would just delete them all at the same time. [P7]

All participants (10/10, 100%) reported that the Headspace app was easy to use. While the majority (9/10, 90%) liked the *minimalistic and cuteness* of the graphics, 1 parent expressed a strong dislike of the narrator's voice:

> They're simple and kind of cute. [P10]

> They were fun and pretty generic, kind of cute. [P7]

> The voice of the narrator. Not relaxing to be told what to do by a white dude. [P8]

None of the activities emerged as most or least liked. Half of the participants were aware that Headspace had activities for children. While 40% (4/10) thought that using the app would

help their child with stress management, only 1 (10%) parent tried the app with their 3-year-old child and reported that it was not a good experience. At the end of the study, 50% (5/10) reported decreases in stress, 40% (4/10) reported no change in stress levels, and 10% (1/10) reported an increase in stress levels:

> *At the end of the study, I found out I can do mindfulness exercises on my own without the app. The app taught me some tricks which I use a lot.* [P10]

Discussion

Principal Findings

We conducted a feasibility and acceptability study to begin to better understand when, how, and under what conditions a digital health intervention could support parental stress management in real time. The findings of this research demonstrate several areas for refinement before conducting a full-scale efficacy trial. First, in contrast to traditional mindfulness intervention recruitment efforts through mental health centers or community or school settings [20], we were able to meet our screening benchmark solely through efforts taken on the web and word of mouth, which did not include paid web-based advertisements (eg, Facebook advertisements). However, we lost nearly 67% (32/48) of eligible participants during enrollment, just shy of our enrollment benchmark (n=16; the benchmark was 20). This conversion rate is consistent with in-person, family-based health promotion programs, despite our intervention being offered without in-person or telehealth meetings. Parent interest in practical solutions for stress management was reaffirmed by the number of participants screened and eligible. However, subpar enrollment and onboarding (n=10; the benchmark was 16 or 100%) rates highlighted competing demands for parent attention, the need for simplicity in everyday solutions, and the ability to integrate into family routines. For those who completed onboarding, the 30-day intervention delivery was viewed as a success, based on 8 of 10 (80%) participants completing the intervention as intended (the benchmark was 80%) and the 72% EMA response rate (the benchmark was 60%). However, it is important to note that this subset may reflect parents who were highly motivated and technically savvy and may not be generalizable to a larger sample. Furthermore, our sample was predominantly mothers, similar to prior evidence [20], and lacked racial or ethnic diversity (the benchmark was 24%). Recruitment efforts may be strengthened by diversifying recruitment strategies to include both free and paid web-based advertisements, as well as traditional partnerships with the community, schools, and clinics.

Another limitation of the feasibility trial was technology acceptance. Familiarity with the Google Play Store and Headspace brand recognition facilitated onboarding and were viewed as helpful among this sample of parents. While leveraging commercial brands and academic-industry partnerships in research may be beneficial, it is not always possible. Owing to other overwhelming requests for partnership, Headspace was unable to commit to a formal partnership at the time of our request. The formative nature of the proposed work, limited research budget (eg, startup funds), and timing (desire

to align the project implementation with the school calendar when family routines are more stable) likely limited any additional opportunities for partnership. Therefore, we used research-developed apps to capture smartphone paradata that would otherwise be proprietary to gain insight into real-world settings. In addition to passively capturing when and for how long the Headspace app was open, App Logger provided objective smartphone paradata that may have been otherwise overlooked, such as if a smartphone was set to silent when a prompt was delivered. However, technology acceptance of research-developed apps was impacted by concerns about privacy and familiarity, similar to other research [73,74]. The most notable challenge during onboarding was downloading an app from outside the Google Play Store (eg, GitHub). Unfamiliarity with installing apps from a source other than the Google Play Store contributed to inquiries from 50% (8/16) of the enrolled parents. Reliance on only Android devices is a common limitation of many mobile health studies that use open science research-developed apps. The Android operating system allows third-party apps to sample from more sensors and system smartphone logs than apps running on the Apple operating system. Efficient and cost-effective solutions in cross-platform mobile development are needed by industry (eg, Google and Apple) to help streamline resources available to researchers [75,76]. Conversely, in lieu of familiarity or brand recognition, family health researchers need to identify strategies for data collection that reduce participant burden (eg, passive data collection) but are viewed as acceptable.

Owing to variability in the literature, no benchmarks have been established for prompt verbiage, timing, or frequency. We deemed prompt wording acceptable in this population (9/10, 90%). However, both prompt frequency and timing were deemed unacceptable owing to subpar ratings (6/10, 60%), despite parents having control over prompt timing and prompt frequency subject to randomization. In exit surveys, most parents indicated that the prompts came at inconvenient times and were often viewed as reminders. Most dismissed the prompt at least once over 30 days (8/10, 80%). If a prompt was dismissed, most parents reported that it was because they were too busy. Only 1 of 10 (10%) parents reported that the study prompt was one of the *tons of other notifications* and would just delete them all. We found a lag of approximately 185 minutes from prompt delivery to action (eg, either launch Headspace app or dismiss notifications), which is congruent with other studies that found a delay of approximately 163 minutes between notification and action [51]. We observed the time-of-day variability for when the Headspace app was opened, noting that some participants tended to use the Headspace app around the same time each day, more so than others. However, there was insufficient data to investigate whether participants habituated to using the Headspace app during a particular time or in response to a prompt. This hypothesis is worth investigating in future trials to further inform habit formation. The literature is mixed on whether it is beneficial to give participants control over when a prompt is delivered. Prior research shows that both user-designated times to receive a prompt and giving the user control of prompt timing failed to enhance the use of a smartphone stress management app [77]. Intelligent prompts guided by sensor-driven machine learning algorithms that adapt

to the user's context may be beneficial for increasing user engagement [77-79]. It has also been suggested that tools to unobtrusively gauge and manage day-to-day stress may be improved by considering contextual information [34,40,80]. However, other research has shown that after 20 days of receiving machine learning suggestions, participants favored self-selecting their intervention, potentially seeking novelty after the algorithm became too locked in or limited in offerings [78]. However, little has been done to understand the nuanced context of everyday family life. We found that while most parents reported using a calendar to facilitate work and family schedules, family calendars captured the exceptions to everyday family routines (eg, appointments or practice) but did not account for daily family activities such as dinnertime, laundry, homework, crying babies, etc. A needs assessment conducted with a larger, more diverse sample and within the context of everyday family life may help to not only determine opportunities for stress support but may also help to identify how much support is needed based on stress severity.

Across all participants, we observed collectively higher engagement with the app on days when a prompt was sent and for a longer duration compared with days in which a prompt was not sent. However, engagement with apps decreased over time. This parallels other research suggesting that prompts impact near-time engagement with the app [30,77] but are not sustained over time [29,43,77,78]. We did not observe a descriptively higher engagement with any mindfulness activity on days when a prompt was sent. We observed more instances in which mindfulness activities were supported by the app versus without the app. Unfortunately, the Beehive app was only able to timestamp when a prompt was sent and unable to capture and record participant-indicated time frames for when to receive the prompt. On the basis of when prompts were sent, we observed peaks during the late morning and night hours. These findings are consistent with findings from a recent review of objective user engagement with 93 mental health apps [81]. Baumel et al [81] found that mindfulness apps were among the most used, with patterns of use exhibiting two peaks, in the morning (7 AM-9 AM) and at night (10 PM-midnight). We also observed that engagement with Headspace was higher in weeks 1 and 2 than in weeks 3 and 4. By weeks 3 and 4, we observed a potential trade-off, where trends in both self-reported and passively collected data indicated a decrease in app usage, while parent-reported mindfulness without the app increased. These findings suggest that mHealth interventions may help parents to learn or practice stress-regulating skills, to be further practiced or applied independent of the app.

In this study, we objectively observed app usage and self-reported interactions with the Headspace app. Most observed interactions with Headspace were ≤60 seconds in duration, while self-reported interactions averaged 5 to 10 minutes per day. This apparent discrepancy between the observed and reported values may reflect a few different scenarios. Unfortunately, a third EMA option to report using both strategies (mindfulness with the app and on their own) was overlooked during development, which may have contributed to some of the discrepancies. Findings may show a response bias, in which parents responded in a manner perceived as desirable. Alternatively, the discrepancy may reflect parents' perceived time actually spent engaging with the app, such as on days when a prompt was sent. Finally, the prompts were capable of launching the Headspace app. Observations may reflect instances where parents intended to dismiss the prompt but accidentally launched and then closed the app. However, this finding is not necessarily a discrepancy and might reflect the self-reported aggregation of interim engagements throughout the day.

The findings suggest that parents may benefit from flexible interventions, allowing freedom to learn the necessary skills and practice or apply them during more resource-limited times. This is further supported by our findings that although app usage decreased over time, the use of learned skills increased over time. On the basis of this observation, disengagement with the app is not necessarily a negative finding. As suggested in the literature, sufficient versus sustained engagement may be a more useful gauge of engagement [34,41]. More research is needed to understand how much exposure to intervention content is necessary to support stress regulation. Despite the brief duration of app engagement, parents in this study reported that Headspace was easy to use and helpful in managing stress. Although the evidence favors a reduction in negative affect following a mindfulness intervention [22,82,83], we did not observe any patterns or trends. In a pilot study by Champion et al [26], participants who used Headspace showed improvements in self-reported life satisfaction, resilience, and perceived stress after 10 days, with medium to large effect sizes found in self-reported life satisfaction, resilience, and perceived stress after 30 days of use. However, participants in that study reported engaging with the app for at least 10 minutes per session and averaged 6.2 days use out of the first 10 days [26]. Other mHealth stress research examined the effects of microinterventions (eg, using a text prompt that instructs the user what to do along with a link to support the skill) [78]. They found benefits for interventions lasting >60 seconds, but those benefits diminished with usage time greater than approximately 200 seconds [78]. Morrison et al [77] found the average participant log-in to a smartphone stress app to be 4 minutes (240 seconds). They concluded that the timing and frequency of push notifications may increase the exposure to intervention content, but not necessarily use [77]. Prior research suggests that a few days of quick log-ins may be sufficient to enable users to learn new skills to practice without continued guidance from the app [41,47,77]. Breathing exercise apps have one of the lowest 30-day retention rates relative to apps that incorporate mindfulness, trackers, and peer support [81]. However, mindfulness apps often design guided meditations for repeated use [81]. Headspace activities favored by parents in this study focused on breathing, tips on how to make mindfulness a part of their day, and restlessness and sleep. Another strategy to consider in future work is to use a stealth health approach, where the intervention is woven into existing routines and the target (eg, stress management) is a side effect but not a primary motivator for participation [47,84].

Limitations and Future Directions

This was a pilot MRT feasibility study designed to address the following question: *Can it work?* [58] Despite a myriad of

recruitment strategies, it was challenging to recruit parents. Participants self-identified as parents who experienced stress and were interested in learning more about how mindfulness might help manage stress. Parenting stress, as evidenced during the pandemic, comes with unique challenges that differentiate this population from the general public. Our sample was predominantly White and female. This is similar to data on the characteristics of paid subscribers to the Calm app, another consumer-based mindfulness app [27]. Calm app users are also predominantly female (8778/10,981, 79.94%). It was difficult to recruit men or fathers compared with women or mothers and diverse populations, which is similar to several other family-based studies [47,85,86]. However, we strongly recommend conducting a needs assessment among diverse groups in future research efforts to include fathers, low-income populations, and racial or ethnic minorities who are often underrepresented in this literature. Finally, more research is warranted among pediatric populations, given the interrelated health between the parent and child. We found that parents did not use or find the Headspace kid pack useful. This differs from recent findings among a sample of parents who use the Calm app [87]. Just over half of the parents who use Calm (1537/2944, 52.21%) reported that older children used Calm to reduce stress, whereas younger children used Calm to improve sleep. A recent review of preventive digital mental health interventions for children and young people (n=21 interventions) found a need for more (1) research among children aged ≤10 years, (2) co-design processes with children, and (3) research among children from vulnerable and underserved backgrounds [88].

Conclusions

Preliminary findings suggest that parents are receptive to using mindfulness apps to support stress management, and prompts are likely to increase engagement with the app. However, we identified several implementation challenges in the current trial, specifically the need to optimize prompt timing and frequency as a strategy to engage users in preventive digital health. Commercially available mindfulness apps appear acceptable among this sample of parents and may provide an opportunity to expose parents to mindfulness skills that may be later practiced without the app. More research is warranted to understand how much time is necessary to spend using a stress management app for parents to learn the necessary skills and translate those skills into everyday life. As the field of mHealth evolves, strategies to engage users in preventive digital mental health interventions must also evolve to increase the likelihood of intervention success.

Authors' Contributions

LM, MS, and DAA wrote the original manuscript. MS, LM, FO, and INS designed the study. LM served as the lead principal investigator and led recruitment, study implementation, and data collection efforts. FO and MS designed and implemented research-developed mobile apps for this study. DAA, MS, and LM conducted all the statistical analyses. All authors edited and reviewed the final manuscript.

Conflicts of Interest

Headspace gifted the study with 10 30-day trial subscriptions free of charge. DAA is co-employed by United Health Group outside of the submitted work.

Multimedia Appendix 1
Descriptive demographics.
[DOCX File , 17 KB - formative_v6i3e30606_app1.docx]

Multimedia Appendix 2
Push notification text.
[DOCX File , 15 KB - formative_v6i3e30606_app2.docx]

References

1. Ingram R, Luxton D. Vulnerability-stress models. In: Development of Psychopathology: A Vulnerability-stress Perspective. Thousand Oaks: SAGE Publications, Inc; 2005:32-46.
2. Deater-Deckard K. Parenting Stress. New Haven, CT: Yale University Press; 2014:9-89.
3. Durtschi JA, Soloski KL, Kimmes J. The dyadic effects of supportive coparenting and parental stress on relationship quality across the transition to parenthood. J Marital Fam Ther 2017 Apr;43(2):308-321. [doi: 10.1111/jmft.12194] [Medline: 27701778]
4. Hardy DF, Power TG, Jaedicke S. Examining the relation of parenting to children's coping with everyday stress. Child Dev 1993 Dec;64(6):1829-1841. [doi: 10.1111/j.1467-8624.1993.tb04216.x] [Medline: 8112121]

5. Burgdorf V, Szabó M, Abbott MJ. The effect of mindfulness interventions for parents on parenting stress and youth psychological outcomes: a systematic review and meta-analysis. Front Psychol 2019;10:1336 [FREE Full text] [doi: 10.3389/fpsyg.2019.01336] [Medline: 31244732]

6. Townshend K, Jordan Z, Stephenson M, Tsey K. The effectiveness of mindful parenting programs in promoting parents' and children's wellbeing: a systematic review. JBI Database System Rev Implement Rep 2016 Mar;14(3):139-180. [doi: 10.11124/JBISRIR-2016-2314] [Medline: 27532143]

7. Ouellette SJ, Russell E, Kryski KR, Sheikh HI, Singh SM, Koren G, et al. Hair cortisol concentrations in higher- and lower-stress mother-daughter dyads: a pilot study of associations and moderators. Dev Psychobiol 2015 Jul;57(5):519-534. [doi: 10.1002/dev.21302] [Medline: 25820649]

8. Herman JP, McKlveen JM, Ghosal S, Kopp B, Wulsin A, Makinson R, et al. Regulation of the hypothalamic-pituitary-adrenocortical stress response. Compr Physiol 2016 Mar 15;6(2):603-621 [FREE Full text] [doi: 10.1002/cphy.c150015] [Medline: 27065163]

9. Vliegenthart J, Noppe G, van Rossum EF, Koper JW, Raat H, van den Akker EL. Socioeconomic status in children is associated with hair cortisol levels as a biological measure of chronic stress. Psychoneuroendocrinology 2016 Mar;65:9-14. [doi: 10.1016/j.psyneuen.2015.11.022] [Medline: 26708067]

10. Lee J, Papa F, Jaini P, Alpini S, Kenny T. An epigenetics-based, lifestyle medicine-driven approach to stress management for primary patient care: implications for medical education. Am J Lifestyle Med 2020;14(3):294-303 [FREE Full text] [doi: 10.1177/1559827619847436] [Medline: 32477032]

11. Terlizzi E, Zablotsky B. Mental Health Treatment Among Adults: United States, 2019. Hyattsville, MD: National Center for Health Statistics; 2020:1-8.

12. Nerurkar A, Bitton A, Davis RB, Phillips RS, Yeh G. When physicians counsel about stress: results of a national study. JAMA Intern Med 2013 Jan 14;173(1):76-77 [FREE Full text] [doi: 10.1001/2013.jamainternmed.480] [Medline: 23403892]

13. Beck J. Cognitive Behavior Therapy, Second Edition: Basics and Beyond. New York, United States: Guilford Press; 2011:1-16.

14. Sharma M, Rush SE. Mindfulness-based stress reduction as a stress management intervention for healthy individuals: a systematic review. J Evid Based Complementary Altern Med 2014 Oct;19(4):271-286. [doi: 10.1177/2156587214543143] [Medline: 25053754]

15. Nielsen L, Riddle M, King JW, NIH Science of Behavior Change Implementation Team, Aklin WM, Chen W, et al. The nih science of behavior change program: transforming the science through a focus on mechanisms of change. Behav Res Ther 2018 Feb;101:3-11 [FREE Full text] [doi: 10.1016/j.brat.2017.07.002] [Medline: 29110885]

16. Duncan LG, Coatsworth JD, Greenberg MT. A model of mindful parenting: implications for parent-child relationships and prevention research. Clin Child Fam Psychol Rev 2009 Sep;12(3):255-270 [FREE Full text] [doi: 10.1007/s10567-009-0046-3] [Medline: 19412664]

17. Birtwell K, Williams K, van Marwijk H, Armitage CJ, Sheffield D. An exploration of formal and informal mindfulness practice and associations with wellbeing. Mindfulness (N Y) 2019;10(1):89-99 [FREE Full text] [doi: 10.1007/s12671-018-0951-y] [Medline: 30662573]

18. Parsons CE, Madsen MA, Jensen KL, Kæseler S, Fjorback LO, Piet J, et al. Smartphone monitoring of participants' engagement with home practice during mindfulness-based stress reduction: observational study. JMIR Ment Health 2020 Jan 14;7(1):e14467 [FREE Full text] [doi: 10.2196/14467] [Medline: 31934873]

19. Ruskin D, Young M, Sugar C, Nofech-Mozes J. Mindfulness and acceptance interventions for parents of children and adolescents diagnosed with chronic medical conditions: a systematic review. J Altern Complement Med 2021 Feb;27(2):120-135. [doi: 10.1089/acm.2020.0191] [Medline: 33151744]

20. Xie Q, Dai X, Lyu R, Lu S. Effects of mindfulness-based parallel-group interventions on family functioning and child and parent mental health: a systematic review and meta-analysis. Mindfulness 2021 Aug 24;12(12):2843-2864. [doi: 10.1007/s12671-021-01728-z]

21. Parenting approaches and concerns. Pew Research Center. URL: https://www.pewresearch.org/social-trends/2015/12/17/3-parenting-approaches-and-concerns/ [accessed 2022-02-16]

22. Schumer MC, Lindsay EK, Creswell JD. Brief mindfulness training for negative affectivity: a systematic review and meta-analysis. J Consult Clin Psychol 2018 Jul;86(7):569-583 [FREE Full text] [doi: 10.1037/ccp0000324] [Medline: 29939051]

23. Wu R, Liu L, Zhu H, Su W, Cao Z, Zhong S, et al. Brief mindfulness meditation improves emotion processing. Front Neurosci 2019;13:1074 [FREE Full text] [doi: 10.3389/fnins.2019.01074] [Medline: 31649501]

24. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR Mhealth Uhealth 2015 Aug 19;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]

25. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the mindfulness meditation mobile app "Calm" to reduce stress among college students: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jun 25;7(6):e14273 [FREE Full text] [doi: 10.2196/14273] [Medline: 31237569]

26. Champion L, Economides M, Chandler C. The efficacy of a brief app-based mindfulness intervention on psychosocial outcomes in healthy adults: a pilot randomised controlled trial. PLoS One 2018;13(12):e0209482 [FREE Full text] [doi: 10.1371/journal.pone.0209482] [Medline: 30596696]

27. Huberty J, Vranceanu A, Carney C, Breus M, Gordon M, Puzia ME. Characteristics and usage patterns among 12,151 paid subscribers of the calm meditation app: cross-sectional survey. JMIR Mhealth Uhealth 2019 Nov 03;7(11):e15648 [FREE Full text] [doi: 10.2196/15648] [Medline: 31682582]

28. Rayan A, Ahmad M. Mindfulness and parenting distress among parents of children with disabilities: a literature review. Perspect Psychiatr Care 2018 Apr;54(2):324-330. [doi: 10.1111/ppc.12217] [Medline: 28247409]

29. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

30. Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, et al. To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. JMIR Mhealth Uhealth 2018 Nov 29;6(11):e10123 [FREE Full text] [doi: 10.2196/10123] [Medline: 30497999]

31. Ben-Zeev D, Schueller SM, Begale M, Duffecy J, Kane JM, Mohr DC. Strategies for mHealth research: lessons from 3 mobile intervention studies. Adm Policy Ment Health 2015 Mar;42(2):157-167 [FREE Full text] [doi: 10.1007/s10488-014-0556-2] [Medline: 24824311]

32. Guendelman S, Medeiros S, Rampes H. Mindfulness and emotion regulation: insights from neurobiological, psychological, and clinical studies. Front Psychol 2017;8:220 [FREE Full text] [doi: 10.3389/fpsyg.2017.00220] [Medline: 28321194]

33. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 08;13:23-47 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]

34. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]

35. Syme S, King A. Interventions to promote health and prevent disease: perspectives on clinical trials past, present, and future. In: New Directions in the Sociology of Aging. Washington, D.C., United States: The National Academies Press; 2013:284-300.

36. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]

37. Menon V, Rajan TM, Sarkar S. Psychotherapeutic applications of mobile phone-based technologies: a systematic review of current research and trends. Indian J Psychol Med 2017;39(1):4-11 [FREE Full text] [doi: 10.4103/0253-7176.198956] [Medline: 28250552]

38. Régnier F, Chauvel L. Digital inequalities in the use of self-tracking diet and fitness apps: interview study on the influence of social, economic, and cultural factors. JMIR Mhealth Uhealth 2018 Apr 20;6(4):e101 [FREE Full text] [doi: 10.2196/mhealth.9189] [Medline: 29678807]

39. Hoffmann A, Christmann CA, Bleser G. Gamification in stress management apps: a critical app review. JMIR Serious Games 2017 Jun 07;5(2):e13 [FREE Full text] [doi: 10.2196/games.7216] [Medline: 28592397]

40. Mohr DC, Riper H, Schueller SM. A solution-focused research approach to achieve an implementable revolution in digital mental health. JAMA Psychiatry 2018 Feb 01;75(2):113-114. [doi: 10.1001/jamapsychiatry.2017.3838] [Medline: 29238805]

41. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]

42. Pfammatter AF, Marchese SH, Pellegrini C, Daly E, Davidson M, Spring B. Using the preparation phase of the multiphase optimization strategy to develop a messaging component for weight loss: formative and pilot research. JMIR Form Res 2020 May 13;4(5):e16297 [FREE Full text] [doi: 10.2196/16297] [Medline: 32347804]

43. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 08;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]

44. Bidargaddi N, Pituch T, Maaieh H, Short C, Strecher V. Predicting which type of push notification content motivates users to engage in a self-monitoring app. Prev Med Rep 2018 Sep;11:267-273 [FREE Full text] [doi: 10.1016/j.pmedr.2018.07.004] [Medline: 30109172]

45. Rothman A, Bartels R, Wlaschin J, Salovey P. The strategic use of gain- and loss-framed messages to promote healthy behavior: how theory can inform practice. J Commun 2006 Aug 4;56(suppl_1):S202-S220. [doi: 10.1111/j.1460-2466.2006.00290.x]

46. Walton A, Nahum-Shani I, Crosby L, Klasnja P, Murphy S. Optimizing digital integrated care via micro-randomized trials. Clin Pharmacol Ther 2018 Jul;104(1):53-58 [FREE Full text] [doi: 10.1002/cpt.1079] [Medline: 29604043]

47. Militello L, Melnyk BM, Hekler EB, Small L, Jacobson D. Automated behavioral text messaging and face-to-face intervention for parents of overweight or obese preschool children: results from a pilot study. JMIR Mhealth Uhealth 2016 Mar 14;4(1):e21 [FREE Full text] [doi: 10.2196/mhealth.4398] [Medline: 26976387]

48. Sharifi M, Dryden EM, Horan CM, Price S, Marshall R, Hacker K, et al. Leveraging text messaging and mobile technology to support pediatric obesity-related behavior change: a qualitative study using parent focus groups and interviews. J Med Internet Res 2013 Dec 06;15(12):e272 [FREE Full text] [doi: 10.2196/jmir.2780] [Medline: 24317406]

49. Muench F, Baumel A. More than a text message: dismantling digital triggers to curate behavior change in patient-centered health interventions. J Med Internet Res 2017 May 26;19(5):e147 [FREE Full text] [doi: 10.2196/jmir.7463] [Medline: 28550001]

50. Hsieh C, Tangmunarunkit H, Alquaddoomi F, Jenkins J, Kang J, Ketcham C, et al. Lifestreams: a modular sense-making toolset for identifying important patterns from everyday life. In: Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems. 2013 Presented at: Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems; Nov 11 - 15, 2013; Roma Italy. [doi: 10.1145/2517351.2517368]

51. Sobolev M. Digital nudging: using technology to nudge for good. SSRN. 2021. URL: https://papers.ssrn.com/abstract=3889831 [accessed 2022-02-16]

52. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]

53. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. Health Psychol 2015 Dec;34S:1220-1228 [FREE Full text] [doi: 10.1037/hea0000305] [Medline: 26651463]

54. Okeke F, Sobolev M, Estrin D. Towards a framework for mobile behavior change research. In: Proceedings of the Technology, Mind, and Society. 2018 Presented at: TechMindSociety '18: Proceedings of the Technology, Mind, and Society; Apr 5 - 7, 2018; Washington DC USA. [doi: 10.1145/3183654.3183706]

55. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]

56. Headspace homepage. Headspace. URL: https://www.headspace.com/ [accessed 2022-02-17]

57. Okeke F, Sobolev M, Dell N, Estrin D. Good vibrations: can a digital nudge reduce digital overload? In: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services. 2018 Presented at: MobileHCI '18: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services; Sep 3 - 6, 2018; Barcelona Spain. [doi: 10.1145/3229434.3229463]

58. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]

59. Goldberg LR. The structure of phenotypic personality traits. Am Psychol 1993 Jan;48(1):26-34. [doi: 10.1037//0003-066x.48.1.26] [Medline: 8427480]

60. John O, Srivastava S. The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In: Handbook of Personality: Theory and Research. New York, United States: The Guilford Press; 1999:102-131.

61. Jeronimus BF, Riese H, Sanderman R, Ormel J. Mutual reinforcement between neuroticism and life experiences: a five-wave, 16-year study to test reciprocal causation. J Pers Soc Psychol 2014 Oct;107(4):751-764. [doi: 10.1037/a0037009] [Medline: 25111305]

62. Karsten J, Penninx BW, Riese H, Ormel J, Nolen WA, Hartman CA. The state effect of depressive and anxiety disorders on big five personality traits. J Psychiatr Res 2012 May;46(5):644-650 [FREE Full text] [doi: 10.1016/j.jpsychires.2012.01.024] [Medline: 22349302]

63. Ormel J, Jeronimus BF, Kotov R, Riese H, Bos EH, Hankin B, et al. Neuroticism and common mental disorders: meaning and utility of a complex relationship. Clin Psychol Rev 2013 Jul;33(5):686-697 [FREE Full text] [doi: 10.1016/j.cpr.2013.04.003] [Medline: 23702592]

64. Crawford JR, Henry JD. The Depression Anxiety Stress Scales (DASS): normative data and latent structure in a large non-clinical sample. Br J Clin Psychol 2003 Jun;42(Pt 2):111-131. [doi: 10.1348/014466503321903544] [Medline: 12828802]

65. Clarke K, Cooper P, Creswell C. The parental overprotection scale: associations with child and parental anxiety. J Affect Disord 2013 Nov;151(2):618-624 [FREE Full text] [doi: 10.1016/j.jad.2013.07.007] [Medline: 23916305]

66. Crawford PB, Gosliner W, Anderson C, Strode P, Becerra-Jones Y, Samuels S, et al. Counseling Latina mothers of preschool children about weight issues: suggestions for a new framework. J Am Diet Assoc 2004 Mar;104(3):387-394. [doi: 10.1016/j.jada.2003.12.018] [Medline: 14993861]

67. Shata ZN, Abu-Nazel MW, Fahmy SI, El-Dawaiaty AA. Efficacy of a psychosocial intervention for parents of children with attention deficit hyperactivity disorder, Alexandria, Egypt. J Egypt Public Health Assoc 2014 Apr;89(1):9-15. [doi: 10.1097/01.EPX.0000443989.38424.d4] [Medline: 24717395]

68. Pollak J, Adams P, Gay G. PAM: a photographic affect meter for frequent, in situ measurement of affect. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI '11: Proceedings of the

SIGCHI Conference on Human Factors in Computing Systems; May 7 - 12, 2011; Vancouver BC Canada. [doi: 10.1145/1978942.1979047]

69. Sobolev M, Vitale R, Wen H, Kizer J, Leeman R, Pollak JP, et al. The Digital Marshmallow Test (DMT) diagnostic and monitoring mobile health app for impulsive behavior: development and validation study. JMIR Mhealth Uhealth 2021 Jan 22;9(1):e25018. [doi: 10.2196/25018]

70. Brooke J. SUS: a 'quick and dirty' usability scale. In: Usability Evaluation In Industry. Boca Raton, Florida, United States: CRC Press; 1996:189-194.

71. Uses and misuses of pilot studies. NIH National Center for Complementary and Integrative Health. URL: https://www.nccih.nih.gov/research/blog/uses-and-misuses-of-pilot-studies [accessed 2022-02-17]

72. Militello L, Sezgin E, Huang Y, Lin S. Delivering perinatal health information via a voice interactive app (SMILE): mixed methods feasibility study. JMIR Form Res 2021 Mar 01;5(3):e18240 [FREE Full text] [doi: 10.2196/18240] [Medline: 33646136]

73. Schueller SM, Neary M, O'Loughlin K, Adkins EC. Discovery of and interest in health apps among those with mental health needs: survey and focus group study. J Med Internet Res 2018 Jun 11;20(6):e10141 [FREE Full text] [doi: 10.2196/10141] [Medline: 29891468]

74. Lipschitz J, Miller CJ, Hogan TP, Burdick KE, Lippin-Foster R, Simon SR, et al. Adoption of mobile apps for depression and anxiety: cross-sectional survey study on patient interest and barriers to engagement. JMIR Ment Health 2019 Jan 25;6(1):e11334 [FREE Full text] [doi: 10.2196/11334] [Medline: 30681968]

75. Harari GM, Lane ND, Wang R, Crosier BS, Campbell AT, Gosling SD. Using smartphones to collect behavioral data in psychological science: opportunities, practical considerations, and challenges. Perspect Psychol Sci 2016 Nov;11(6):838-854 [FREE Full text] [doi: 10.1177/1745691616650285] [Medline: 27899727]

76. Biørn-Hansen A, Grønli T, Ghinea G, Alouneh S. An empirical study of cross-platform mobile development in industry. Wireless Commun Mobile Comput 2019 Jan 03;2019:1-12 [FREE Full text] [doi: 10.1155/2019/5743892]

77. Morrison LG, Hargood C, Pejovic V, Geraghty AW, Lloyd S, Goodman N, et al. Correction: the effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: an exploratory trial. PLoS One 2018;13(5):e0198008 [FREE Full text] [doi: 10.1371/journal.pone.0198008] [Medline: 29787600]

78. Paredes P, Gilad-Bachrach R, Hernandez J, Czerwinski M, Roseway A. PopTherapy: coping with stress through pop-culture. In: Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare. 2014 Presented at: 8th International Conference on Pervasive Computing Technologies for Healthcare; May 20 - 23, 2014; Oldenburg Germany. [doi: 10.4108/icst.pervasivehealth.2014.255070]

79. Klasnja P, Smith S, Seewald NJ, Lee A, Hall K, Luers B, et al. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of HeartSteps. Ann Behav Med 2019 May 03;53(6):573-582 [FREE Full text] [doi: 10.1093/abm/kay067] [Medline: 30192907]

80. Hao T, Rogers J, Chang H, Ball M, Walter K, Sun S, et al. Towards precision stress management: design and evaluation of a practical wearable sensing system for monitoring everyday stress. iproc 2017 Sep 22;3(1):e15 [FREE Full text] [doi: 10.2196/iproc.8441]

81. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]

82. Gotink RA, Hermans KS, Geschwind N, De Nooij R, De Groot WT, Speckens AE. Mindfulness and mood stimulate each other in an upward spiral: a mindful walking intervention using experience sampling. Mindfulness (N Y) 2016;7(5):1114-1122 [FREE Full text] [doi: 10.1007/s12671-016-0550-8] [Medline: 27642373]

83. Nisbet EK, Zelenski JM, Grandpierre Z. Mindfulness in nature enhances connectedness and mood. Ecopsychology 2019 Jun;11(2):81-91. [doi: 10.1089/eco.2018.0061]

84. Wearing JR, Nollen N, Befort C, Davis AM, Agemy CK. iPhone app adherence to expert-recommended guidelines for pediatric obesity prevention. Child Obes 2014 Apr;10(2):132-144 [FREE Full text] [doi: 10.1089/chi.2013.0084] [Medline: 24655230]

85. Chacko A, Fabiano GA, Doctoroff GL, Fortson B. Engaging fathers in effective parenting for preschool children using shared book reading: a randomized controlled trial. J Clin Child Adolesc Psychol 2018;47(1):79-93 [FREE Full text] [doi: 10.1080/15374416.2016.1266648] [Medline: 28103110]

86. Skelton JA, Martin S, Irby MB. Satisfaction and attrition in paediatric weight management. Clin Obes 2016 Apr;6(2):143-153. [doi: 10.1111/cob.12138] [Medline: 27008068]

87. Puzia M, Laird B, Green J, Huberty J. Parents' perceptions of their children's engagement in a consumer-based meditation mobile app: cross-sectional survey study. JMIR Pediatr Parent 2020 Nov 13;3(2):e24536. [doi: 10.2196/24536]

88. Bergin AD, Vallejos EP, Davies EB, Daley D, Ford T, Harold G, et al. Preventive digital mental health interventions for children and young people: a review of the design and reporting of research. NPJ Digit Med 2020 Oct 15;3(1):133 [FREE Full text] [doi: 10.1038/s41746-020-00339-7] [Medline: 33083568]

Abbreviations

DASS-21: Depression Anxiety Stress Scale 21-item
EMA: ecological momentary assessment
IRB: Institutional Review Board
JITAI: just-in-time adaptive intervention
mHealth: mobile health
MRT: microrandomized trial
PAM: Photographic Affect Meter
PI: principal investigator

XSL•FO
RenderX

Original Paper

Egyptian Students Open to Digital Mental Health Care: Cross-Sectional Survey

Mostafa Mamdouh[1,2], MD; Andy Man Yeung Tai[1], BSc; Jean Nicolas Westenberg[1], BSc; Farhud Shams[1], BS; Kerry Jang[1], PhD; Adel Badawy[2], MD, PhD; Houssam Elsawi[2], MD, PhD; Michael Krausz[1], MD, PhD

[1]Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada
[2]Department of Psychiatry, Tanta University, Tanta, Egypt

Corresponding Author:
Andy Man Yeung Tai, BSc
Department of Psychiatry
University of British Columbia
Vancouver Campus 430-5950
University Boulevard, David Strangway Building
Vancouver, BC, V6T 1Z3
Canada
Phone: 1 6474024088
Email: andytai7@student.ubc.ca

Abstract

Background: In Egypt, the shortage of mental health services, particularly for adolescents and young adults, is apparent. Electronic mental health (EMH) has been proposed as a solution to bridge the gap and better address the needs of young people. However, EMH is new to Egypt and its acceptability among target populations is crucial to its implementation and success.

Objective: The objective of this study is to explore the interest of Egyptian youth in EMH, identify perceived barriers to EMH, and inform the design of EMH tools to best address the needs of youth.

Methods: A web-based cross-sectional survey was distributed among medical students at Tanta University in Egypt. Chi-square and one-way analysis of variance tests were performed for inferential analyses using a significance level of .05.

Results: Of the 707 individuals who completed the survey (90.9% response rate), 60.5% (428) were female, 62% (438) lived in urban and suburban areas, and the mean age of the sample was 20.5 (SD 1.8) years. The vast majority of participants (522/707, 73.8%) had already used the internet to find information about mental health problems, but the information was unsatisfactory for about half of them (386/707, 54.6%). Almost all students reported that they would prefer web-based therapy if EMH were available through a trustworthy national web-based platform for youth mental health (601/707, 85%). Students believed that emotional difficulties, social support, and coping strategies were the main topics that EMH should help with. The most common perceived barriers for EMH use in Egypt were concerns about privacy (382/707, 54%) and a lack of technology literacy and unfamiliarity with EMH (352/707, 50%).

Conclusions: EMH is a promising strategy for addressing gaps in the mental health care for young people. To construct and implement a digital system of care that addresses the unique needs and preferences of youth, adolescents and young adults should be involved in the co-development and design.

Introduction

Globally, mental illness is the most common health problem among postsecondary students [1]. Moreover, without access to appropriate treatment services, mental health problems often worsen throughout an individual's adult life [2]. It is well established that mental health services are extremely scarce in low- and middle-income countries, estimated to be 200 times lower than that in high-income countries [3]. Youth in low-income countries who struggle with mental illness therefore

represent a particularly vulnerable population [1,4]. The aim of this study is to explore electronic mental health (EMH) as a promising strategy to address gaps in the mental health care of young people in the low-income world, by specifically interacting and engaging with students in Egypt.

Though half of the Egyptian population is under the age of 25 years, health care structures are not well equipped to handle mental illness in youth [5]. In a national study of 13,000 high school students in 3 different regions of Egypt, 20%-30% of students had mental health problems, with similar rates reported in several other studies [6]. In Egyptian universities, the prevalence of mental illness is even higher, with up to 37% of undergraduate students fulfilling the criteria for moderate depression [7]. However, there are only 1.44 psychiatrists and 0.11 psychologists per 100,000 people in Egypt [8], and the shortage of mental health services is particularly apparent for youth [5]. Within the government's mental health workforce, only roughly 3% work in child and adolescent services [9]. Moreover, only around 1% of health care professionals who work in schools are trained in mental health [8].

To provide care, improve access, and build capacity, web-based resources have become increasingly necessary tools, especially during the COVID-19 pandemic [10,11]. EMH interventions have proven effective in providing mental health services to different vulnerable youth populations and have been increasingly used across the world [12]. EMH domains include web-based resources for mental health, such as information, risk assessments, professional and peer counselling, group therapy, cognitive behavioral therapy, and telepsychiatry, that address the plethora of mental illness comorbidities. Implementation of these tools is an iterative process that includes co-development for specific youth populations. Successful EMH platforms for youth mental health services have been developed in different jurisdictions, including Australia, Ireland, and the United Kingdom, with varying designs and methods [12,13]. Such technology-based interventions are also more accessible and cost-effective [14]. According to recent evidence, many Arabs would rather use EMH services than visit a mental health provider, which would help them overcome some preexisting barriers such as stigma, cost, and physical distance [15]. Interestingly, the use of EMH in conjunction with medication-assisted treatment has proven more effective than standalone treatment for substance use disorders in youth [16]. A major strength of web-based interventions is the ability to provide youth with a continuum of care, starting from early diagnosis to continued peer and professional support [17].

EMH is especially relevant to the Egyptian health care system given the high burden of mental health problems, low access to available resources, and high internet accessibility and mobile phone ownership [18]. Existing studies have investigated the implementation and use of platforms such as electronic medical records in Egypt, but no studies have evaluated how youth perceive these technologies [19]. Psychiatrists in Egypt have agreed that EMH could be the solution to building necessary capacity for youth mental health care, especially given the scarcity of resources for this demographic and familiarity with technology among youth [20]. Despite the ubiquity of smartphones and internet access in Egypt and in other Arab

countries, EMH is largely an untapped resource [21,22]. Attempts have already been made using Arabic mobile apps for depression and anxiety, but clear gaps are evident given the low quality, lack of engagement, and absence of evidence-based resources [23]. Crucial to the development and implementation of EMH in Egypt is the opinion of target populations, namely adolescents and young adults. Little is known about the types of mental health information sought by young people on the internet, how EMH can address their needs, and their perceived barriers to using EMH.

The objectives of this study are to gauge the level of interest of Egyptian youth in EMH approaches, highlight features of web-based interventions that are most appealing to students, and identify the perceived barriers to using EMH in this population. Responses were analyzed by gender and living region to best determine the mental health needs of particular target groups among youth. The findings of this study aim to help inform the development and design of EMH tools that best address mental health issues among youth not only in Egypt but also in other Arabic countries. More broadly, this formative research is an integral part of program development, as it explores the feasibility and acceptability of EMH among Egyptian youth before large-scale summative evaluations such as randomized controlled trials.

Methods

Survey Design

The survey instrument was initially developed by researchers and health care providers from Tanta University, Tanta, Egypt, and the University of British Columbia, Vancouver, Canada, as part of an ongoing collaboration regarding student mental health. Following its design, 11 students from Tanta University (6 females and 5 males; mean age 20.4 years) were invited to participate in a preliminary analysis of the survey instrument in two 1.5-hour sessions. After revising the instrument based on comments from the students in the workshops, the survey was piloted with a sample of 60 students from Tanta University. Some questions were again modified based on student feedback to improve the overall quality of the instrument. The final version of the instrument was then distributed within the student body of the Faculty of Medicine at Tanta University. Medical school students from Tanta University were selected as the target sample because of their high health literacy level and their ability to address broad issues related to health and well-being and to understand the culture, etiquette, and customs of the general Egyptian public [24].

The survey consisted of 28 questions and included multiple choice questions, dichotomous questions, and Likert scales. The survey was in English since medical education in Egypt is also delivered in English. The survey consisted of both quantitative and qualitative questions meant to explore patterns of internet use and EMH strategies, thereby giving youth a platform to voice their attitudes toward and perceptions of EMH.

Recruitment

The survey was distributed within the entire student body in October 2019, thereby including students from all 6 years of

medical school. Participants were recruited by nonrandom sampling via the student council. The student council of the Faculty of Medicine at Tanta University was provided with a hyperlink to the consent form and questionnaire; the council then sent this link to all students within the student body. When respondents clicked on the link, they were first asked if they were interested in participating in the study. Only those who ticked "yes" could access the survey. Participants were reminded that they could withdraw at any time without giving any justifications and without any negative consequences. Participants were permitted to skip any question they were unwilling to answer and could change any answer over the course of the entire survey. Participation was voluntary and students did not receive any reimbursement.

The following description of EMH was provided to the participants: "E-Mental Health refers to the delivery of mental health services (treatment, information and support) via the internet or mobile phone. This can be through websites, web applications, video conferencing, chat or email. Some of these services, such as video conferencing or online counselling, involve direct one on one contact with a mental health professional. Other e-mental health services, such as web applications or information websites, involve less or no contact with mental health professionals" [25].

Data Collection and Analysis

Responses were collected electronically using the Qualtrics platform. To eliminate the possibility of duplicated responses, students were allowed to access the survey only once. All data gathered by this study are confidential and anonymous. Chi-square and one-way analysis of variance tests were performed for inferential analyses using a significance level of .05. Descriptive and inferential statistical analyses were executed using SPSS 25 (IBM Corp, 2017) [26]. The results from the web-based survey have been reported according to the Checklist for Reporting Results of Internet E-Surveys [27].

Ethical Considerations

The study received approval from the Tanta University Ethics Board (31674/07/17). All participants provided informed consent.

Results

Participant Demographics and General Internet Use

The web-based survey link was opened by 778 Egyptian medical school students, of which 707 consented and completed the survey (90.9% response rate). More than half the number of participants were female (428/707, 60.5%), and the mean age was 20.5 (SD 1.8) years (Table 1). All academic years, from first to sixth, were represented in our sample. The majority of participants were living in urban and suburban areas (438/707, 62%), while only 38% (269/707) were living in rural areas (Table 1).

When asked about their internet use, almost all the participants (652/707, 92.2%) said that they used the internet several times a day. Participants mainly accessed the internet using their smartphone (689/707, 97.5%) and mostly from their homes (692/707, 97.8%). The main reasons for using the internet included communication and social media (622/707, 88%) as well as information gathering (611/707, 86.4%; Table 1).

Table 1. Participant demographics and internet use (N=707).

Demographics	Value, n (%)
Gender	
Female	428 (60.5)
Male	279 (39.5)
Age (years)	
18	66 (9.3)
19	215 (30.4)
20	116 (16.4)
21	82 (11.6)
22	125 (17.7)
23	64 (9)
24	15 (21)
25	24 (3.4)
Region	
Urban and suburban	438 (62)
Rural	269 (38.1)
Frequency of internet use	
Several times per week	23 (3.3)
Once per day	32 (4.5)
Several times per day	652 (92.2)
Device	
Computer	247 (34.9)
Smartphone	689 (97.5)
Another device	67 (9.5)
Place of internet use	
Home	602 (97.9)
University	351 (49.7)
Public areas or transport	306 (43.3)
Other	96 (13.6)
Main reasons for internet use	
Information (content)	611 (86.4)
Support (community)	215 (30.4)
Communication (social media)	622 (88)
Electronic commerce (e-commerce)	52 (7.4)
Gaming	276 (39)
Others	156 (22)

Internet Use for Health Information

The vast majority of participants used the internet to find information about physical and mental health problems (595/707, 84.2% and 522/707, 73.8%, respectively), but the information found on the internet was satisfactory for only about half the number of participants (386/707, 54.6%; Multimedia Appendix 1). Use of the internet to find information about mental health problems was not significantly different between genders and living regions, but significantly more women than men used the internet to find information on physical health problems (373/428, 87.2% vs 222/279, 79.6%; P=.007; Multimedia Appendix 1; Table 1). Significantly more individuals from rural areas preferred Arabic to English, in terms of the language of information on the internet (P=.001).

Participants revealed that the main reasons for using the internet as a source of web-based mental health help was convenience, user-friendliness, and privacy (696/707, 95.6%, 604/707, 85%, and 594/707, 84%, respectively; Figure 1).

Figure 1. Advantages of web-based interventions. Participants were asked whether they agreed with, disagreed with, or were neutral (neither agreed nor disagreed) toward various advantages of web-based interventions.
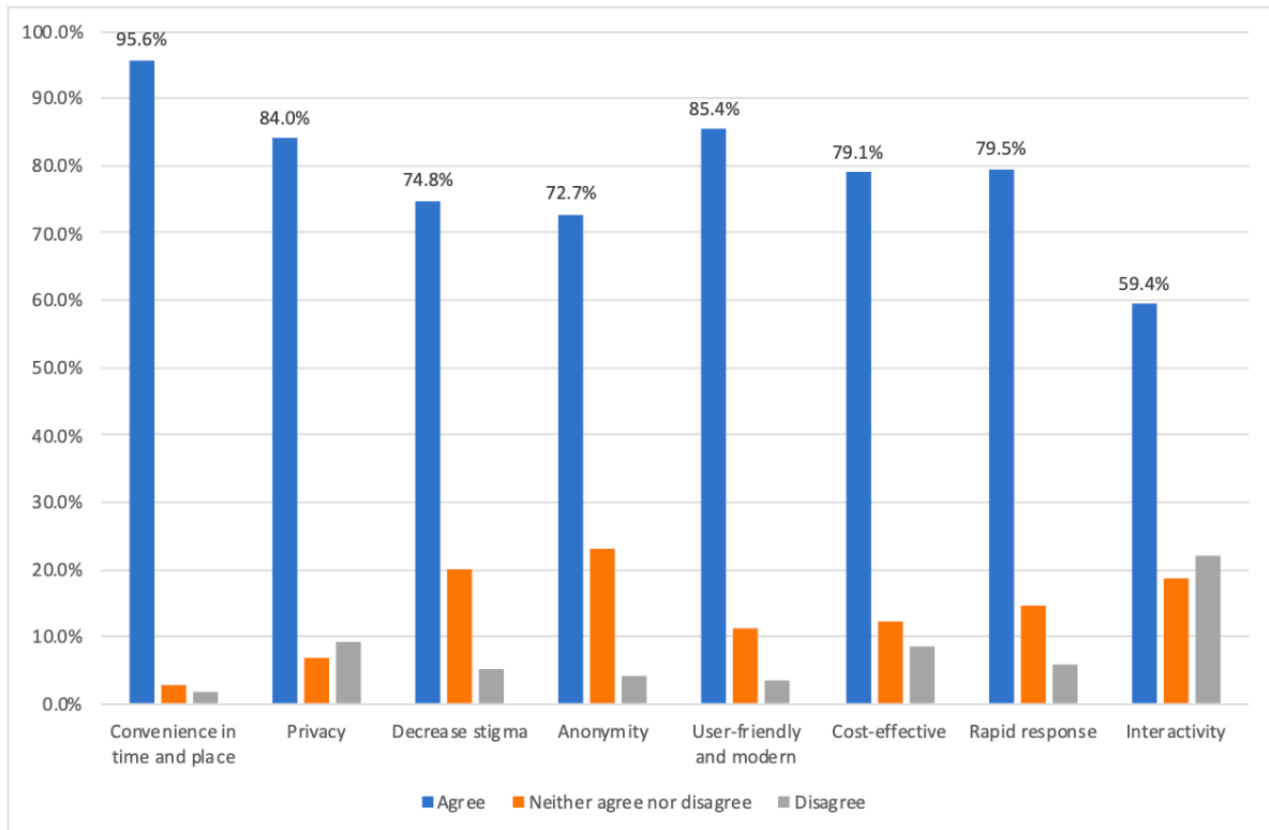
Knowledge of and Interest in EMH

Over half the number of participants (388/707, 54.9%) did not know that mental health websites and mobile apps existed, and about half the number of participants (359/707, 50.8%) said that web-based mental health services would be an attractive option for them (Table 2). About half the number of participants (366/707, 51.8%) said that they would prefer web-based therapy to conventional psychotherapy. Almost all students (601/707, 85%) reported that they preferred web-based therapy if EMH were available through a trustworthy national web-based platform for youth mental health (Table 2).

Table 2. Knowledge of and interest in web-based mental health services (N=707).

Characteristic	Value, n (%)
Knowledge of preexisting mental health websites and applications	
Yes	319 (45.1)
No	388 (54.9)
Web-based mental health services as an attractive option	
Yes	359 (50.8)
No	107 (15.2)
I don't know	241 (34.1)
Preference for web -based therapy over conventional psychotherapy	
Yes	366 (51.8)
No	341 (48.2)
Preference for web-based therapy if available through a trustworthy national platform	
Yes	601 (85)
No	106 (15)

Knowledge about EMH services was significantly different between genders and living regions (*P*<.001 and *P*<.007, respectively; Multimedia Appendix 1; Table 1). Males (149/428, 53.4%) and individuals living in urban areas (215/438, 49.1%) knew more about existing mental health websites and apps than females (170/279; 39.7%) and individuals living in rural areas (104/269, 38.7%), respectively. There were no significant differences in terms of interest in EMH between genders and living regions (Multimedia Appendix 1; Table 1).

Priorities for and Barriers to EMH Development

Participants believed that emotional difficulties, social support, dealing with stressors, and coping strategies were the main topics that EMH should help with (Table 3). When asked about how participants wanted the information on an EMH platform to be delivered, most participants suggested videos explaining mental health topics (539/707, 76.2%); skills training for improving coping strategies, time management, and self-care (381/707, 53.9%); and web-based mood and behavior assessments (351/707, 49.6%). The most common perceived barriers to EMH use were concerns about confidentiality and privacy issues (382/707, 54%), uncertainty toward and unfamiliarity with EMH (352/707, 50%), and technical difficulties (242/707, 34%).

Table 3. Priorities for and barriers to electronic mental health development (N=707).

Characteristic	Value, n (%)
Priorities for EMH[a]	
Emotional difficulties	431 (60.1)
Social support	424 (60)
Dealing with stressors	420 (59.4)
Learning coping strategies	403 (57)
Mental health	365 (51.6)
Sexual education	240 (34)
Physical well-being	230 (32.5)
Self-harm behaviors	218 (30.8)
Cultural- and religious-sensitive topics	207 (29.3)
Substance use	148 (21)
Others	52 (7.4)
Information displayed on EMH	
Videos to help explain mental health topics	539 (76.2)
Skills training modules	381 (53.9)
Web-based tools to assess mood and behavior	351 (49.7)
Self-guided web-based intervention	312 (44.1)
Pictures to help explain mental health topics	298 (42.2)
Family involvement and support	248 (35.1)
Web-based peer connection	167 (23.6)
Information delivered in game format	147 (20.8)
Other	40 (5.7)
Barriers to EMH use	
Privacy issues and confidentiality	382 (54)
Technical issues and difficulties	242 (34.2)
Cost	166 (23.5)
Validity and reliability	191 (27)
Uncertainty toward or unfamiliarity with EMH	352 (49.8)

[a]EMH: electronic mental health.

There were many significant differences between genders and living regions (Multimedia Appendix 2; Table 2). More women thought that web-based platforms should help in dealing with stressors (*P*=.03) and self-harm behaviors (*P*=0.03), whereas more men were interested in seeing web-based platforms help with coping strategies (*P*<.001), sexual education (*P*<.001), and

substance use (*P*=.02). Individuals living in urban areas reported coping strategies (*P*=.01), sexual education (*P*=.001), and cultural- and religious-sensitive topics (*P*=.003) as priorities.

Discussion

EMH

Among Egyptian university students between 18 and 25 years of age, the vast majority report using the internet mainly for social media and general information gathering, which are findings similar to those of previously reported surveys in the same age group from different countries such as the United States, England, and Spain [28-30]. Moreover, Egyptian youth use the internet to find information about mental health, which was also seen among European students [31]. The internet is certainly a popular source of information for youth around the world and seems to be an accessible medium for mental health information.

In Egypt, there is high demand for EMH services that are safe and reliable, and which may be well addressed by a nationally supported platform. A notable example of such a trustworthy and nationally implemented EMH program is eheadspace in Australia, which provides web-based and phone support services to vulnerable youth [32]. eheadspace includes 3 different tools, including group chats, web-based interactive dashboards, and private chats. Group chats increase communication and better enable web-based discussions facilitated by mental health professionals between students and like-minded individuals. The dashboard allows vulnerable youth to collect and manage resources that they find helpful for their daily routines [32]. Private chat sessions accompanied by health professionals can provide youth with intimate conversations that help meet their own individual needs [33]. Since its launch in 2012, eheadspace has created a fully operational eMental health digital ecosystem with an eMental health portal that provides funded services to Australian youth. eheadspace has proven to be an effective solution and has been a key part of the infrastructure addressing mental health among Australian youth [34]. Countries in the Middle East could follow suit to create effective and much needed EMH platforms across the region.

Web-based therapy relative to conventional therapy would be preferred for many Egyptian students if they were diagnosed as having mental illness. This is consistent with findings of an Iranian study showing that students were more willing to use EMH services if available [35]. Conversely, a majority of Irish university students reported preferring face-to-face support rather than web-based support for a mental health problem because it was more reliable and allowed better communication [25]. This discrepancy between youth of different countries, mainly European and Middle Eastern countries, is likely due to several different factors such as stigma around mental health and accessibility to in-person mental health services. Indeed, the scarcity of face-to-face services and resources in Egypt could push students toward web-based solutions if those are most available to them. Students in Egypt reported accessibility and convenience among the main reasons for preferring EMH as a solution for mental health support along with other important benefits such as anonymity and mitigating stigma. More

generally, individuals with mental health issues tend to be less resistant to the use of EMH programs if these are effective, and they report willingness to pursue services if these are made easily accessible to them [36].

Differences in EMH Needs

As expected, women and men reported different perceived needs and opinions regarding web-based services. For instance, women seemed more interested than men in searching the internet for physical health problems possibly because of a higher prevalence of stigma experienced by women with mental illness that leads to the somatization of psychological symptoms, which is common in Arab countries [37]. In terms of priorities for EMH, women expressed more interest in interventions that could help address self-harm behaviors and that could help deal with stressors. Conversely, men seemed to want EMH interventions to provide education on coping behaviors, sexual education, and substance use. Moreover, individuals living in rural and urban areas also reported varying needs. For instance, rural populations seemed to prefer the language to be Arabic on the internet, whereas urban populations preferred English. This highlights the importance of considering language in the provision of web-based resources. The different needs of the various subpopulations demonstrate the importance of providing mental health information that is as individualized as possible. In general, it seems that the relationship between patients and physicians has become less hierarchical and more client-provider oriented [38]. EMH solutions can address a wide variety of issues in this regard and offer functionalities to tailor content to the needs of the individual, for example, by gauging a user's need through in-app assessments [39].

Barriers to EMH

The most important perceived barrier for Egyptian students was confidentiality and privacy of personal information. This is in agreement with the results of a web-based survey that examined consumer expectations and potential challenges of EMH services across several countries such as Australia, Iran, the Philippines, and South Africa [35]. Interestingly, the highest rates of participant willingness to try EMH services are seen in low-income countries such as Iran and Egypt, which also reported a high number of barriers and concerns with patient confidentiality and the protection of personal information [35]. It is important for program developers to recognize the functional and technical assistance that individuals may need to use such services [35]. In high-income countries, there seems to be less concern with how data are handled, possibly because of past positive experiences in dealing with sensitive information on already developed web-based health platforms [40]. Another common concern among Egyptian students was their unfamiliarity with the range of functionalities in EMH as well as technical approaches. These findings were similar to those of another cross-sectional study of university students in the United Arab Emirates, in which half the number of participants had never heard of mobile mental health care apps and 75% had never used these kinds of apps even during the COVID-19 pandemic [41]. Considering that the survey sample consisted of medical school students, familiarity is likely even lower in the general young population and thus the barriers to EMH use

could be greater. These findings highlight the need for increased discussions around the use of EMH in the Middle East. EMH should be a topic in the curriculum for all Middle Eastern students to increase their comfort with it, thereby furthering its implementation and acceptability. Increasing the presence of EMH resources in the everyday lives of youth would also allow for better co-design and collaboration between developers and students. This would help address concerns on the validity and reliability of the available content, which was also highlighted as a perceived barrier by the youth in our sample [42].

Limitations

The findings are not completely generalizable to the general young population because the sample comprised well-educated and technologically savvy students. Moreover, the questionnaire was web-based, and all participants had to use the internet to complete the questionnaire, which demonstrated at least some degree of digital literacy. It must also be considered that all students at this particular university are expected to use the internet as a way of staying up to date on courses. Future studies should examine perceptions about EMH in other subpopulations in Egypt, such as in university students from other specialties (not health care related) and youth not enrolled in postsecondary educational programs.

Conclusions and Implications

The findings of this study highlight that young people attending university are active users of the internet and are willing to use the internet for mental health information and support. EMH is therefore a feasible strategy for addressing gaps in the mental health care for young people in Egypt. Moreover, target populations (eg, males vs females, urban vs rural) as well as topics of interests (eg, self-harm, substance use) must be considered when implementing web-based solutions, given the different needs and preferences of these populations as outlined by this study. Finally, EMH platforms should prioritize informational videos and skills training modules as a way to display content while also addressing privacy and confidentiality issues, which were identified as barriers to EMH use among Egyptian students.

Based on the results of this survey, a virtual mental health clinic is being developed for the students at Tanta University, which will be the first web-based and evidence-based intervention designed specifically for university students in Egypt. The clinic, which will be developed in collaboration with international and local experts, will provide support to students and begin to address gaps in the mental health care for young people in Egypt. Nevertheless, further studies in this area are needed to better understand the feasibility of EMH in the broader Egyptian population and in other Arabic countries. Researchers and mental health clinicians in the Middle East must work together with the youth in their countries to develop and implement web-based interventions that are accepted and used by this population. Students should strive to develop and use EMH platforms to provide compassion and love for the generations to come.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Knowledge of and interest in electronic mental health services by gender and region.
[DOCX File , 21 KB - formative_v6i3e31727_app1.docx]

Multimedia Appendix 2
Perceived priorities of electronic mental health among Egyptian students by gender and region.
[DOCX File , 19 KB - formative_v6i3e31727_app2.docx]

References

1. Wiljer D, Shi J, Lo B, Sanches M, Hollenberg E, Johnson A, et al. Effects of a mobile and web app (Thought Spot) on mental health help-seeking among college and university students: randomized controlled trial. J Med Internet Res 2020 Oct 30;22(10):e20790. [doi: 10.2196/20790] [Medline: 33124984]
2. Keyes CLM. Mental illness and/or mental health? Investigating axioms of the complete state model of health. J Consult Clin Psychol 2005 Jun;73(3):539-548. [doi: 10.1037/0022-006X.73.3.539] [Medline: 15982151]
3. Saxena S, Thornicroft G, Knapp M, Whiteford H. Resources for mental health: scarcity, inequity, and inefficiency. Lancet 2007 Sep 08;370(9590):878-889. [doi: 10.1016/S0140-6736(07)61239-2] [Medline: 17804062]
4. Rickwood D, Thomas. Conceptual measurement framework for help-seeking for mental health problems. Psychol Res Behav Manag 2012;5:173-183 [FREE Full text] [doi: 10.2147/PRBM.S38707] [Medline: 23248576]

5. Saeed S. Aims of Egypt: Assessment of Governmental Mental Health System Egypt (2016-2017). RUN| Communities and collections | NOVA Medical School - Faculty of Medical Sciences (NMS-FCM). 2018 Jun 25. URL: https://run.unl.pt/handle/10362/40174 [accessed 2022-03-14]

6. Rabie M, Shaker N, Gaber E, El-Habiby M, Ismail D, El-Gaafary M, et al. Prevalence updates of substance use among Egyptian adolescents. Middle East Curr Psychiatry 2020 Feb 05;27(1) [FREE Full text] [doi: 10.1186/s43045-019-0013-8]

7. Ibrahim AK, Kelly SJ, Glazebrook C. Analysis of an Egyptian study on the socioeconomic distribution of depressive symptoms among undergraduates. Soc Psychiatry Psychiatr Epidemiol 2012 Jun;47(6):927-937. [doi: 10.1007/s00127-011-0400-x] [Medline: 21626055]

8. Hussein H, Shaker N, El-Sheikh M, Ramy HA. Pathways to child mental health services among patients in an urban clinical setting in Egypt. Psychiatr Serv 2012 Dec;63(12):1225-1230. [doi: 10.1176/appi.ps.201200039] [Medline: 23070112]

9. Mental health atlas 2017: resources for mental health in the Eastern Mediterranean Region. Geneva: World Health Organization. Regional Office for the Eastern Mediterranean; 2019:116.

10. Krausz M, Westenberg JN, Vigo D, Spence RT, Ramsey D. Emergency response to COVID-19 in Canada: platform development and implementation for eHealth in crisis management. JMIR Public Health Surveill 2020 May 15;6(2):e18995 [FREE Full text] [doi: 10.2196/18995] [Medline: 32401218]

11. Naslund JA, Aschbrenner KA, Araya R, Marsch LA, Unützer J, Patel V, et al. Digital technology for treating and preventing mental disorders in low-income and middle-income countries: a narrative review of the literature. Lancet Psychiatry 2017 Jun;4(6):486-500 [FREE Full text] [doi: 10.1016/S2215-0366(17)30096-2] [Medline: 28433615]

12. Schellenberg M, Hatcher S, Thapliyal A. E-Mental Health in Canada: Transforming the Mental Health System Using Technology. Mental Health Commission of Canada. 2014. URL: https://www.mentalhealthcommission.ca/wp-content/uploads/drupal/MHCC_E-Mental_Health-Briefing_Document_ENG_0.pdf [accessed 2022-03-14]

13. McGorry PD, Goldstone SD, Parker AG, Rickwood DJ, Hickie IB. Cultures for mental health care of young people: an Australian blueprint for reform. Lancet Psychiatry 2014 Dec;1(7):559-568. [doi: 10.1016/S2215-0366(14)00082-0] [Medline: 26361315]

14. Lal S, Adair CE. E-mental health: a rapid review of the literature. Psychiatr Serv 2014 Jan 01;65(1):24-32. [doi: 10.1176/appi.ps.201300009] [Medline: 24081188]

15. Kayrouz R, Dear BF, Karin E, Fogliati VJ, Gandy M, Keyrouz L, et al. Acceptability of mental health services for anxiety and depression in an Arab sample. Community Ment Health J 2018 Aug;54(6):875-883. [doi: 10.1007/s10597-018-0235-y] [Medline: 29368132]

16. Marsch LA, Guarino H, Acosta M, Aponte-Melendez Y, Cleland C, Grabinski M, et al. Web-based behavioral treatment for substance use disorders as a partial replacement of standard methadone maintenance treatment. J Subst Abuse Treat 2014 Jan;46(1):43-51 [FREE Full text] [doi: 10.1016/j.jsat.2013.08.012] [Medline: 24060350]

17. Krausz M, Ward J, Ramsey D. From Telehealth to an Interactive Virtual Clinic. In: e-Mental Health. Switzerland: Springer, Cham; 2016.

18. Digital 2020: Egypt. DATAREPORTAL. URL: https://datareportal.com/reports/digital-2020-egypt [accessed 2021-01-04]

19. Badran MF. eHealth in Egypt: The demand-side perspective of implementing electronic health records. Telecomm Policy 2019 Jul;43(6):576-594. [doi: 10.1016/j.telpol.2019.01.003]

20. Kamel MM, Westenberg JN, Choi F, Tabi K, Badawy A, Ramy H, et al. Electronic mental health as an option for Egyptian psychiatry: cross-sectional study. JMIR Ment Health 2020 Aug 13;7(8):e19591 [FREE Full text] [doi: 10.2196/19591] [Medline: 32788155]

21. Salem F. The Arab World Online 2017: Digital Transformations and Societal Trends in the Age of the 4th Industrial Revolution. SSRN | The Arab World Online Policy Report Series. 2017 Oct 26. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3059445 [accessed 2022-03-14]

22. Donker T, Blankers M, Hedman E, Ljótsson B, Petrie K, Christensen H. Economic evaluations of Internet interventions for mental health: a systematic review. Psychol Med 2015 Dec;45(16):3357-3376. [doi: 10.1017/S0033291715001427] [Medline: 26235445]

23. Alhuwail D, Albaj R, Ahmad F, Aldakheel K. The state of mental digi-therapeutics: a systematic assessment of depression and anxiety apps available for Arabic speakers. Int J Med Inform 2020 Mar;135:104056. [doi: 10.1016/j.ijmedinf.2019.104056] [Medline: 31918341]

24. Aluh DO, Anyachebelu OC, Anosike C, Anizoba EL. Mental health literacy: what do Nigerian adolescents know about depression? Int J Ment Health Syst 2018;12:8 [FREE Full text] [doi: 10.1186/s13033-018-0186-2] [Medline: 29467817]

25. Horgan A, Sweeney J. Young students' use of the internet for mental health information and support. J Psychiatr Ment Health Nurs 2010 Mar;17(2):117-123. [doi: 10.1111/j.1365-2850.2009.01497.x] [Medline: 20465756]

26. IBM SPSS Statistics: Version 24. SPSS Software | IBM. 2017. URL: https://www.ibm.com/analytics/spss-statistics-software [accessed 2022-03-14]

27. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]

28. Rogers VL, Griffin MQ, Wykle ML, Fitzpatrick JJ. Internet versus face-to-face therapy: emotional self-disclosure issues for young adults. Issues Ment Health Nurs 2009 Oct;30(10):596-602. [doi: 10.1080/01612840903003520] [Medline: 19742368]

29. Eynon R, Malmberg L. A typology of young people's Internet use: Implications for education. Comput Educ 2011 Apr;56(3):585-595. [doi: 10.1016/j.compedu.2010.09.020]

30. Montagni I, Parizot I, Horgan A, Gonzalez-Caballero J, Almenara-Barrios J, Lagares-Franco C, et al. Spanish students' use of the internet for mental health information and support seeking. Health Informatics J 2016 Jun;22(2):333-354 [FREE Full text] [doi: 10.1177/1460458214556908] [Medline: 25477327]

31. Montagni I, Donisi V, Tedeschi F, Parizot I, Motrico E, Horgan A. Internet use for mental health information and support among European university students: The e-MentH project. Digit Health 2016 Jun 17;2:2055207616653845 [FREE Full text] [doi: 10.1177/2055207616653845] [Medline: 29942558]

32. Rickwood D, Webb M, Kennedy V, Telford N. Who are the young people choosing web-based mental health support? findings from the implementation of Australia's national web-based youth mental health service, eheadspace. JMIR Ment Health 2016 Aug 25;3(3):e40 [FREE Full text] [doi: 10.2196/mental.5988] [Medline: 27562729]

33. McGorry P, Bates T, Birchwood M. Designing youth mental health services for the 21st century: examples from Australia, Ireland and the UK. Br J Psychiatry Suppl 2013 Jan;54:s30-s35. [doi: 10.1192/bjp.bp.112.119214] [Medline: 23288499]

34. Rickwood D, Wallace A, Kennedy V, O'Sullivan S, Telford N, Leicester S. Young people's satisfaction with the online mental health service eheadspace: development and implementation of a service satisfaction measure. JMIR Ment Health 2019 Apr 17;6(4):e12169 [FREE Full text] [doi: 10.2196/12169] [Medline: 30994470]

35. Clough BA, Zarean M, Ruane I, Mateo NJ, Aliyeva TA, Casey LM. Going global: do consumer preferences, attitudes, and barriers to using e-mental health services differ across countries? J Ment Health 2019 Feb;28(1):17-25. [doi: 10.1080/09638237.2017.1370639] [Medline: 28857650]

36. Meurk C, Leung J, Hall W, Head BW, Whiteford H. Establishing and governing e-Mental health care in Australia: a systematic review of challenges and a call for policy-focussed research. J Med Internet Res 2016 Jan 13;18(1):e10 [FREE Full text] [doi: 10.2196/jmir.4827] [Medline: 26764181]

37. Merhej R. Stigma on mental illness in the Arab world: beyond the socio-cultural barriers. IJHRH 2019 Aug 29;12(4):285-298. [doi: 10.1108/ijhrh-03-2019-0025]

38. Boyer CA, Lutfey KE. Examining critical health policy issues within and beyond the clinical encounter: patient-provider relationships and help-seeking behaviors. J Health Soc Behav 2010 Oct 08;51 Suppl(1_suppl):S80-S93. [doi: 10.1177/0022146510383489] [Medline: 20943585]

39. Levin-Zamir D, Bertschi I. Media health literacy, eHealth literacy, and the role of the social environment in context. Int J Environ Res Public Health 2018 Aug 03;15(8):1643 [FREE Full text] [doi: 10.3390/ijerph15081643] [Medline: 30081465]

40. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual Research Review: Digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr 10;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]

41. Drissi N, Alhmoudi A, Al Nuaimi H, Alkhyeli M, Alsalami S, Ouhbi S. Investigating the impact of COVID-19 lockdown on the psychological health of university students and their attitudes toward mobile mental health solutions: two-part questionnaire study. JMIR Form Res 2020 Oct 20;4(10):e19876 [FREE Full text] [doi: 10.2196/19876] [Medline: 32969340]

42. Feijt M, de Kort Y, Bongers I, Bierbooms J, Westerink J, IJsselsteijn W. Mental health care goes online: practitioners' experiences of providing mental health care during the COVID-19 pandemic. Cyberpsychol Behav Soc Netw 2020 Dec 01;23(12):860-864. [doi: 10.1089/cyber.2020.0370] [Medline: 32815742]

Abbreviations

EMH: electronic mental health

Original Paper

Effectiveness of a Smartwatch App in Detecting Induced Falls: Observational Study

Bruce Brew[1,2], MBBS, DMedSci, DSC; Steven G Faux[2,3], MBBS, BA; Elizabeth Blanchard[4], MBA, MSc

[1]Department of Neurology, St Vincent's Hospital, Sydney, Australia

[2]University of New South Wales, Sydney, Australia

[3]Sacred Heart Rehabilitation Service, St Vincent's Hospital, Sydney, Australia

[4]My Medic Watch, Sydney, Australia

Corresponding Author:
Bruce Brew, MBBS, DMedSci, DSC
Department of Neurology
St Vincent's Hospital
390 Victoria Street, Darlinghurst
Sydney, 2010
Australia
Phone: 61 2 8382 4100
Email: Bruce.Brew@svha.org.au

Abstract

Background: Older adults are at an increased risk of falls with the consequent impacts on the health of the individual and health expenditure for the population. Smartwatch apps have been developed to detect a fall, but their sensitivity and specificity have not been subjected to blinded assessment nor have the factors that influence the effectiveness of fall detection been fully identified.

Objective: This study aims to assess accuracy metrics for a novel fall detection smartwatch algorithm.

Methods: We performed a cross-sectional study of 22 healthy adults comparing the detection of induced forward, side (left and right), and backward falls and near falls provided by a smartwatch threshold-based algorithm, with a video record of induced falls serving as the gold standard; a blinded assessor compared the two. Three different smartwatches with two different operating systems were used. There were 226 falls: 64 were backward, 51 forward, 55 left sided, and 56 right sided.

Results: The overall smartwatch app sensitivity for falls was 77%, the specificity was 99%, the false-positive rate was 1.7%, and the false-negative rate was 16.4%. The positive and negative predictive values were 98% and 84%, respectively, while the accuracy was 89%. There were 249 near falls: the sensitivity was 89%, the specificity was 100%, there were no false positives, 11% were false negatives, the positive predictive value was 100%, the false-negative predictive value was 83%, and the accuracy was 93%.

Conclusions: Falls were more likely to be detected if the fall was on the same side as the wrist with the smartwatch. There was a trend toward some smartwatches and operating systems having superior sensitivity, but these did not reach statistical significance. The effectiveness data and modifying factors pertaining to this smartwatch app can serve as a reference point for other similar smartwatch apps.

Introduction

The risk of falling increases with age. Approximately 30% of people older than 65 years and living in the community have a fall at least once a year, with an increase of 5% each year [1]. The incidence is even higher in those living in aged care facilities [2]. This is a major public health problem leading to injuries [1,3], loss of quality of life [1,3], loss of independence [1], placement in assisted-living facilities [4,5] and premature mortality [3]. Fall-related injuries represent 21% of the total health care expenses due to injuries [3] and between 0.85% and 1.5% of the total health care expenditure [6]. Lying on the floor

for a long time after a fall has been associated with serious consequences, with a greater likelihood of hospitalization, decline in activities of daily living, placement into long-term care, and mortality [4,5].

Assistive technologies such as call alarm systems and personal emergency response systems are increasingly available. This also holds true for wearables, defined as devices that can be worn or are in contact with human skin to continuously and closely monitor an individual's activities without interrupting or limiting the users' motions [7]. These are cost-effective in reducing hospital admissions when used within emergency response systems [8,9]. However, these systems are not always used by consumers, in part, due to difficulties activating them, including cognitive impairment at the time of, or prior to, the fall [5].

There is an increasing interest in using sensor systems embedded in smartwatches for health care purposes [10,11]. This is particularly the case with falls detection. Although there are several fall detection devices and apps, none to our knowledge have been subjected to a blinded study to evaluate effectiveness, particularly with a variety of smartwatches and smartphones using different operating systems. This study aims to address these issues.

Methods

Ethics Committee

The procedures followed in this study were conducted according to the principles of the World Medical Association Declaration of Helsinki and were approved by the University of New South Wales and St Vincent's Hospital Human Research Ethics Committee jointly (16/229). The study was independently audited.

Study Design

This is a cross-sectional blinded study comparing the fall detection classification provided by a smartwatch algorithm with a reference standard's classification, in this case, a video record of induced falls.

Participants

A total of 22 volunteer participants deemed to be medically healthy were recruited after satisfying all the inclusion and exclusion criteria. Participants were recruited by distribution of a leaflet on the university campus and compensated for their time. The inclusion criteria were males/females older than 18 years willing and able to provide written informed consent prior to initiation of any study-related procedures. Participants were excluded if they had any of the following: disability that may prevent them from completing the study (eg, severe illness), being suspected of or having a known allergy to any components of the smartwatch, having any injury or medical condition that would be adversely affected by an induced fall, and being pregnant.

Smartwatch Threshold Algorithm

This study used a threshold-based algorithm programmed for different smartwatches. The threshold-based algorithm running on the smartwatch app uses threshold values, or settings, to automatically detect a fall. The frequency of the smartwatch accelerometer is 2 kHz with the algorithm of the app collecting data every 0.01 seconds. The algorithm follows strict rules for the three phases of a fall, as shown in Figure 1. The algorithm was supplied by My Medic Watch.

T1 is defined as the time during which the smartwatch is moving toward the earth (fall time) recording a low acceleration, lower than 1G. T2 is the time during which the smartwatch hits the ground, recording a very high positive acceleration for a short period of time. T3 is the time during which the smartwatch is "almost" immobile on the ground for a long period of time. These threshold values are optimized in the app according to the particular smartwatch and body morphology, including body weight and height. Optimization was performed during the test falls.

A near fall can be recognized when all, or one, of the accelerometer data are close to one of the thresholds, as depicted in Figure 2. We have arbitrarily defined "close" as 20% lower than the fall threshold value.

Figure 1. The threshold-based algorithm settings for fall detection. 1G: force of gravity 9.8 m/s^2; accel: acceleration; T1: time of phase 1 of the fall; T2: time of phase 2 of the fall; T3: time of phase 3 of the fall.
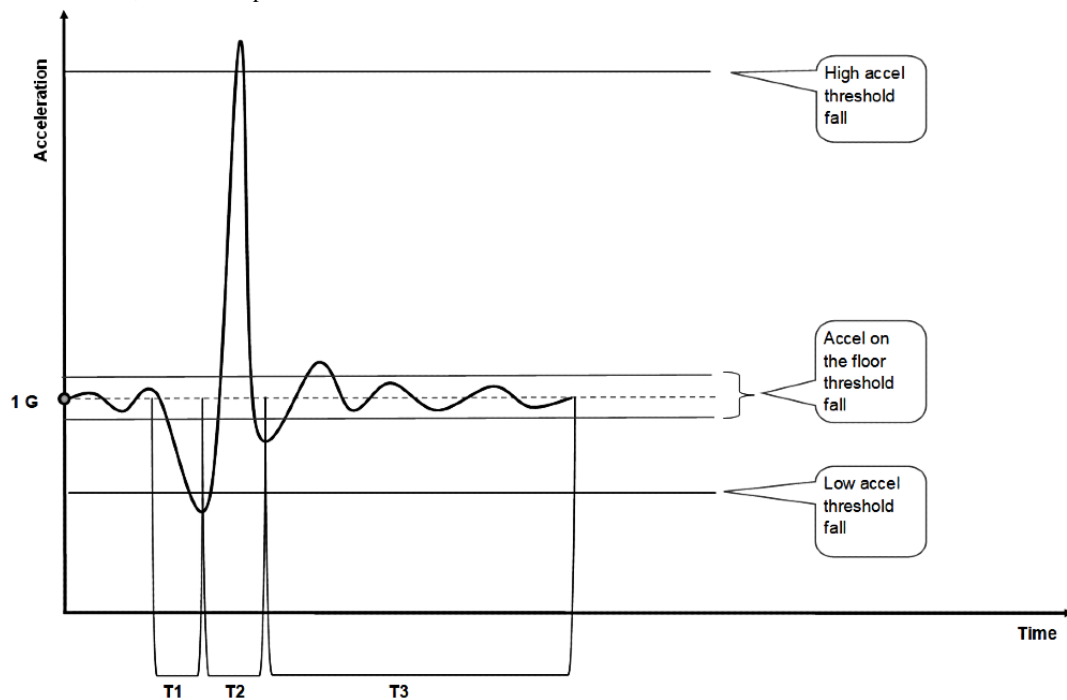
Figure 2. The algorithm threshold settings for the detection of a near fall. 1G: force of gravity 9.8 m/s^2; accel: acceleration; T1: time of phase 1 of the fall; T2: time of phase 2 of the fall; T3: time of phase 3 of the fall.
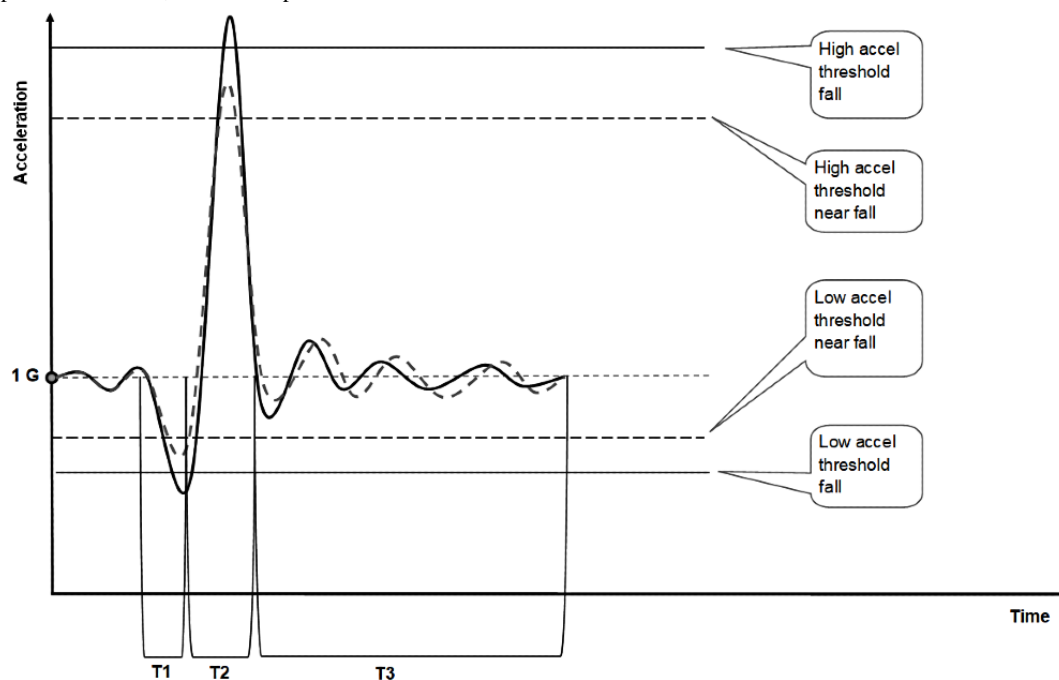
Protocol

Participants were randomly assigned to have either smartwatch model A or model C on one wrist and model B or no device on the other wrist. Model A and C were running one operating system, while model B was running on a different operating system. Every smartwatch contained the fall detection app that was programmed to detect and record falls paired with a smartphone located at the study site. The same app was used for each model. The smartwatches and smartphones used one of two operating systems: android or iOS. Two smartwatches were connected to iOS and one to Android. The versions of iOS and Android were the latest available at the time of the test. The version of the operating system on the smartphone and smartwatch were the same for all participants. The smartphones were linked to the smartwatch (according to the operating system) to communicate stored data of the time-stamped recorded episodes to secure cloud servers that were then compared to the video-recorded events.

Before starting the trial, participants were placed in a crash mat protected area, the smartwatches were placed on the participants' wrists, and a helmet was provided to be used during the tests; no other safety devices were used. Once the trial started, the smartwatch app was set up in monitoring mode and two rounds of four falls were induced in the blindfolded participants. A fall was defined as an event that results in a person coming to rest inadvertently on the ground, floor, or other lower level. A nonfall was defined as any event occurring while both the smartwatch app and the video record were active but excluding a fall or near fall (defined later). In every round, a frontward fall, a right side fall, a left side fall, and a backward fall were induced. These were induced by pushing the participant while standing. The method of fall induction was the same for all participants, executed by the same person. The participants were told of the impending direction of the push. Each assessment took approximately 5 minutes with 8 falls: 2 backward, 2 forward, 2 right, and 2 left. Additionally, up to 3 test falls were performed before the first round to ensure the participants were feeling comfortable with the procedure. Test falls were not included in the analyses. Further, prior to the test falls and between the falls, the participants wore the smartwatches and walked around freely. Near falls where the participant took one or more steps in the direction of the push without falling were also recorded, as there is some evidence that they may presage a fall [12]. This definition is in accord with the traditional definition as applied to this experimental scenario: "a stumble event or loss of balance that would result in a fall if sufficient recovery mechanisms were not activated" [12]. Importantly, the fall-triggering settings were optimized for each participant during the test falls. A non–near fall was defined as any event occurring while both the smartwatch app and the video record were active but excluding a fall or near fall.

During the fall, the algorithm was collecting the acceleration data and the time of the fall. The data collected were in three phases: "prefall" (preparation and walking to the crash mat, several minutes) as soon as the smartwatches were on the participants wrist, "induced fall" (8 falls around 5 minutes), and "post fall," walking back from the crash matt to the area to remove the smartwatches. In addition to this, the falls were recorded by built-in motion-detecting cameras (recording at 50 frames/second) available at the study site, the National Facility for Human Robot Interaction Research, University of New South Wales. Motion detection data were used to indicate when a fall was observed. The video of the falls also contained a timestamp that was used to compare it with the falls detected by the smartwatch app. In this case, the video recorded event was used as a reference standard, and the falls detected by the smartwatches were compared against it.

After all the falls had been induced, the smartwatches and safety equipment were removed, and participants were observed for

approximately 10 minutes: the heart rate, blood pressure, and symptoms (if any) were assessed.

Data Analysis

To perform the analysis of the falls, data were first retrieved from video records of the built-in motion-detecting cameras and coded as a fall or near fall by the authors and a person independent of the conduct of the study. Where there was disagreement, a majority opinion was taken. These were then compared independently by an external person with data retrieved from a fall detection database built to register the falls detected by the smartwatch algorithm. Each fall was classified as a true positive if the smartwatch app detected a fall at the time when the event was recorded on the video, a false positive if the smartwatch detected a fall event that was not recorded on the video, a false negative if the smartwatch did not detect a fall event recorded on the video, and a true negative if neither the smartwatch nor the video recorded a fall. Near falls were similarly analyzed. Results were computed for sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, positive predictive value, negative predictive value, and accuracy. CIs for sensitivity, specificity, and accuracy are "exact" Clopper-Pearson CIs. CIs for the likelihood ratios are calculated using the "Log method." To compare fall and near fall detection by smartwatch model and direction of fall only, sensitivity data were used with chi-square tests and a significance value of $P<.05$. Further data are available on request. Sample size calculations were not formally performed beyond an approximate anticipated number of 20 to 25 participants that could be accommodated for the study given the constraints of the availability of the study site and personnel time.

Results

Characteristics of the Participants and the Falls

A total of 22 participants were enrolled in the study: 14 (63%) females and 8 (36%) males; 20 (91%) completed the whole procedure. Two (9%) females abandoned the study during the process: one after a soft tissue injury and the other for unstated reasons. An average of 7.2 falls was performed for each participant; however, one of the participants withdrew from the study after having performed 5 sets of 8 falls, and another after having performed 1 set of 8 falls. Of the induced 226 falls, 64 were backward, 51 were forward, 55 were left sided, and 56 were right sided. Two participants reported postfall self-limiting symptoms associated with soft tissue injuries, 1 required medication and physiotherapy, and their symptoms resolved after 6 weeks.

Demographic characteristics of the participants are shown in Table 1. With regard to BMI, 1 (6%) female was classified as underweight, 1 male and 1 female were classified as overweight (9%), and 1 (6%) male was classified as obese.

Table 1. Demographic characteristic of the participants.

Gender	Age (years)	Height (cm)	Weight (kg)
Female	25	160	58
Female	24	167	57
Female	24	170	62
Female	28	153	48
Female	21	164	50
Male	18	177	68
Female	19	164	47
Male	19	175	65
Female	19	164	53
Male	21	174	62
Female	25	170	60
Female	24	163	53
Female	23	168	49
Female	18	174	60
Male	18	180	63
Male	38	171	86
Female	33	163	55
Male	23	184	110
Female	45	160	50
Female	52	163	64
Male	32	160	65
Male	42	178	70

Overall Performance of the Algorithm

A total of 12 participants were wearing two smartwatches, model A device on one wrist and model B on the other wrist; 10 participants were wearing only one smartwatch, model C, on one wrist. The overall performances of the algorithm, disregarding the model of the smartwatch, are detailed in Tables 2 and 3. There was no difference in the performance of the algorithm according to which wrist if both were used. Tables 4 and 5 represent the results of near fall detection and the associated statistics. The overall test outcomes are summarized in the following section.

In general, the direction of the fall or near fall did not significantly influence sensitivity. Nonetheless, there was a trend for better detection of backward falls: of the 64 backward falls, 11 were false negatives, giving a sensitivity of 82%, versus forward falls, of which there were 51 with 12 false negatives, giving a sensitivity of 76%. Further, there was a significant difference in fall detection if the fall was to the same side versus opposite side of the wrist that had the smartwatch (left sided and right sided sensitivities combined: 92.5% vs 76.3%; $P=.009$). The same held true for near falls. If the fall was to the same side as the wrist with the smartwatch, there was a 95% sensitivity for left sided falls (55 with 3 false negatives) and 89% sensitivity for right sided falls (56 with 11 false negatives) versus if the fall was on the opposite side as the wrist with the smartwatch, there was 84% sensitivity for left sided falls (55 with 9 false negatives) and 80% sensitivity for right sided falls (56 with 11 false negatives).

Table 2. Fall detection results.

True fall status	Test result, n		Total, n
	Negative (nonfall)	Positive (fall)	
Nonfall	265 (true negative)	3 (false positive 1.7%)	268
Fall	52 (false negative 16.4%)	174 (true positive)	226
Total	317	177	494

Table 3. Statistics for fall detection.

	Value (95% CI)
Sensitivity (%)	76.99 (70.95-82.31)
Specificity (%)	98.88 (96.76-99.77)
Positive likelihood ratio	68.78 (22.27-212.39)
Negative likelihood ratio	0.23 (0.18-0.30)
Positive predictive value (%)	98.31 (94.95-99.44)
Negative predictive value (%)	83.60 (80.05-86.61)
Accuracy (%)	88.87 (85.76-91.50)

Table 4. Near fall detection results.

True near fall status	Test result, n		Total, n
Non–near fall	343 (true negative for all falls, normal falls, and near falls)	0 (false positive)	343
Near fall	43 (false negative when near fall 11.1%)	206 (true positive)	249
Total	386	206	592

Table 5. Statistics for near fall detection.

	Value (95% CI)
Sensitivity (%)	88.86 (85.29-91.82)
Specificity (%)	100 (98.23-100)
Positive likelihood ratio	N/A[a] (no false positives)
Negative likelihood ratio	0.11 (0.08-0.15)
Positive predictive value (%)	100
Negative predictive value (%)	82.73 (78.33-86.39)
Accuracy (%)	92.74 (90.34-94.69)

[a]N/A: not applicable.

Performance by Smartwatch Model

The number of responses for each smartwatch model were A=186, B=186, and C=122. Model A was used 173 times on the left wrist and 13 times on the right wrist. As per Table 6, there were differences among the models according to sensitivity and specificity, but none were significant. This was also true of the operating system. Similar results were found for near falls.

Table 6. Fall detection results by smartwatch models A, B, and C. The direction of the fall did not significantly influence sensitivity in any of the models.

	Value (95% CI)
Model A	
Sensitivity (%)	78.8 (68.6-86.9)
Specificity (%)	99 (94.6-100)
Model B	
Sensitivity (%)	71.8 (61-81)
Specificity (%)	98 (93-99.8)
Model C	
Sensitivity (%)	82.1 (96.6-91.1)
Specificity (%)	100 (94.6-100)

XSL•FO
RenderX

Discussion

The primary goal of this study was to evaluate the validity of an algorithm programmed in commercially available smartwatches to detect induced falls. Our study found that the algorithm had an overall sensitivity of 77% and specificity of 99%. The false-positive rate was very low at 1.7%, while the false-negative rate was 16.4%. The positive and negative predictive values were 98% and 84%, respectively, while the accuracy was 89%. Falls were more likely to be detected if the fall was on the same side as the wrist with the smartwatch. Similar results were found for near falls. There was a trend toward some smartwatches having superior sensitivity, though neither this nor the operating system reached statistical significance.

Several studies have been conducted to assess the performance of wearable devices for fall detection, mostly by using smartphones or other specialized self-created wearable devices [13-15]. However, only a few of these studies have been performed using commercially available smartwatches [16-19]. In addition, this study is the only one to assess the performance of a fall detection algorithm in different commercially available smartwatches with different operating systems using a video recording system as a gold standard and using blinded data analysis.

The fall detection algorithm was threshold based—programmed to send an alert once a predetermined threshold had been breached. Threshold-based algorithms, as opposed to pattern recognition methods, are preferred on smartphone operating systems due to the restrictions on computing and storage capabilities of the devices [16]. Indeed, pattern recognition methods are costly and need massive analyses of data, access to databases, and long training periods.

Casilari and Oviedo-Jimenez [16] tested different algorithms with an LG W110 smartwatch model R, finding that the fall detection performance depends on the algorithm used. However, there were only 4 participants with a total of 40 falls. Sensitivity ranged from 70% to 100% and specificity from 80% to 100% depending on the type of fall. Mauldin et al [18] have studied three different pattern recognition algorithms based on Naive bayes (NB), support vector machine (SVM), and deep learning models by using a Microsoft band 2 smartwatch. In this context, the algorithm tested in our study performed better than their NB and SVM models in sensitivity and precision, and when compared with their deep learning model, our algorithm performed better in precision but not sensitivity. Mauldin et al [18] also declared in their study that they tested an Android wear-based commercially available fall detection app (Rightminder) released on the Google Play store. The sensitivity was only 50%, and no technical details of this app are publicly available.

Further, these studies have used small groups of participants (3-7) performing several falls each (up to 10 per side). From our experience in laboratory settings, the dynamics of the falls are affected by repetition, as participants tend to fall in the same way. We minimized this effect by having a high number of participants (N=22) repeating each fall only twice per side.

Furthermore, the previous studies asked the participants to fall rather than having them fall as a result of being pushed unexpectedly by another person as was done in our study. This approach more accurately reflects a true fall given the spontaneity. The differing protocol designs in these studies make it impossible to accurately compare one against the other.

Our findings suggest that the performance of the algorithm differs among various brand devices. Indeed, the combined performance of brand A and C smartwatches on sensitivity and false-negative rates was higher than the brand B smartwatch. However, the brand B smartwatch precision and thus the false-positive rate is better than brands A and C devices. This is probably related to the differences in the operating systems. Medrano et al [20] explain that in current smartphone operating systems such as Android and iOS, it is difficult to configure specific sampling rates. As the sampling frequencies in both systems are different, the performance of the algorithm will likely be influenced by the operating system used. Moreover, Fudickar et al [21] have investigated the impact of the sampling frequency of the accelerometer on the performance of different threshold-based algorithms in smartphones, concluding that a detection system must deal with the polling frequency of the accelerometer sensors embedded in the device. No studies have been performed regarding this issue on smartwatches; however, it is likely that the situation is the same.

Additionally, our study has found that the performance of the algorithm could be strongly dependent on the smartwatch model. According to Silva et al [22], the performance of a fall detection algorithm could be affected by the quality of the sensors embedded in the device. Additionally, as the manufacturer can change the sensors over time, the performance of the algorithm will also rely on the smartwatch model [16]. This situation could explain the differences we have found between the smartwatch models tested, making it difficult to compare with other studies if they have not used the same smartwatch device and model.

It has been previously reported that the direction of the fall affects the performance of the algorithm used in smartwatches [16,18]. In this context, the performance of the algorithm is largely dependent on which side the fall occurred in relation to the smartwatch. Our algorithm performs better when the fall occurs on the same side of the wrist wearing the smartwatch than when the fall occurs on the opposite side. This is a tendency observed regardless of the smartwatch model. Mauldin et al [18] found a similar performance in the three pattern recognition models they tested. Casilari and Oviedo-Jimenez [16] reported an overall result for side falls; therefore, it is not possible to know if they have found the same tendency.

Regarding the back falls, Mauldin et al [18] found their different algorithm models had poor performance indices in this direction. This was thought to be a consequence of less wrist movement in back falls as compared to other directions of falls. However, our algorithm performed the best on back falls, suggesting that the intensity of the wrist movement or the impact is not affecting the algorithm in this fall direction.

Finally, another factor that could affect the performance of the algorithm in detecting falls in different directions is the participant's body habitus. It has been proposed that height and

weight could affect the performance of the algorithm [23]; thus, implementing personalized settings according to participants' characteristics is a way to improve the algorithm sensitivity. To address these issues of body habitus and smartwatch model, we deliberately adjusted the algorithm settings during the test falls. This likely contributes to the positive results and should be considered in future studies.

Our study has some limitations. First, there was a relatively small number of participants though not in comparison with other published studies. Second, not all participants wore a smartwatch on each arm, potentially influencing the results. However, only 1 participant was wearing one smartwatch; the results were essentially unchanged with that participant's data removed. Third, our participants were healthy in contradistinction to the older adult population who would most likely be using the app. Nonetheless, inducing falls in such participants would expose them to considerable risk.

Despite these reservations, the smartwatch app performed well in comparison to studies of other apps and under more rigorous conditions with more stringent analyses, yielding an accuracy of 89%. Indeed, the field of physical activity sensors generally accepts an accuracy of 70% to 80% [24]. Our future research will be focused on investigating the performance of the algorithm in different smartwatch models by using personalized settings. Moreover, head-to-head studies of fall detection devices in smartwatches using real-world participants and settings are likely to improve available evidence concerning the effectiveness of these devices for consumers such as older adults and regulatory or licensing bodies.

Acknowledgments

Authors' Contributions

BB and SGF contributed to the concept and design of the study. All authors were involved in the implementation of the study and data collection as well as analyses. All authors contributed to the writing of the manuscript. The final version of the paper has been seen and approved by all authors.

Conflicts of Interest

BB reports grants from My Medic Watch during the conduct of the study. In addition, BB has patent AU2017338619 with royalties paid, patent CA3039538Ð with royalties paid, patent CN109843171Ð with royalties paid, patent EP3522782Ð with royalties paid, patent JP2020504806 with royalties paid, patent KR1020190058618 with royalties paid, and patent US20200051688 with royalties paid. BB is a scientific advisor to My Medic Watch Pty Ltd. EB reports grants from My Medic Watch during the conduct of the study. In addition, EB has patent AU2017338619 with royalties paid, patent CA3039538 with royalties paid, patent CN109843171 with royalties paid, patent EP3522782 with royalties paid, patent JP2020504806 with royalties paid, patent KR1020190058618 with royalties paid, and patent US20200051688 with royalties paid. EB is the Director of My Medic Watch Pty Ltd. SGF has nothing to disclose. My Medic Watch provided unrestricted funds to cover the infrastructure costs of the study: ethics submission, research assistant for participant logistics and data collection, and statistician for data analyses.

References

1. Peel NM. Epidemiology of falls in older age. Can J Aging 2011 Mar;30(1):7-19. [doi: 10.1017/S071498081000070X] [Medline: 21401978]

2. Rubenstein LZ. Falls in older people: epidemiology, risk factors and strategies for prevention. Age Ageing 2006 Sep;35 Suppl 2:ii37-ii41. [doi: 10.1093/ageing/afl084] [Medline: 16926202]

3. Hartholt KA, van Beeck EF, Polinder S, van der Velde N, van Lieshout EMM, Panneman MJM, et al. Societal consequences of falls in the older population: injuries, healthcare costs, and long-term reduced quality of life. J Trauma 2011 Sep;71(3):748-753. [doi: 10.1097/TA.0b013e3181f6f5e5] [Medline: 21045738]

4. Tinetti ME, Liu WL, Claus EB. Predictors and prognosis of inability to get up after falls among elderly persons. JAMA 1993 Jan 06;269(1):65-70. [Medline: 8416408]

5. Fleming J, Brayne C, Cambridge City over-75s Cohort (CC75C) study collaboration. Inability to get up after falling, subsequent time on floor, and summoning help: prospective cohort study in people over 90. BMJ 2008 Nov 17;337:a2227 [FREE Full text] [doi: 10.1136/bmj.a2227] [Medline: 19015185]

6. Heinrich S, Rapp K, Rissmann U, Becker C, König HH. Cost of falls in old age: a systematic review. Osteoporos Int 2010 Jun;21(6):891-902. [doi: 10.1007/s00198-009-1100-1] [Medline: 19924496]

7. Gao W, Emaminejad S, Nyein HYY, Challa S, Chen K, Peck A, et al. Fully integrated wearable sensor arrays for multiplexed in situ perspiration analysis. Nature 2016 Jan 28;529(7587):509-514 [FREE Full text] [doi: 10.1038/nature16521] [Medline: 26819044]

8. Roush RE, Teasdale TA, Murphy JN, Kirk MS. Impact of a personal emergency response system on hospital utilization by community-residing elders. South Med J 1995 Sep;88(9):917-922. [doi: 10.1097/00007611-199509000-00006] [Medline: 7660208]

9. Bernstein M. "Low-tech" personal emergency response systems reduce costs and improve outcomes. Manag Care Q 2000;8(1):38-43. [Medline: 11009732]

10. Reeder B, David A. Health at hand: a systematic review of smart watch uses for health and wellness. J Biomed Inform 2016 Oct;63:269-276 [FREE Full text] [doi: 10.1016/j.jbi.2016.09.001] [Medline: 27612974]

11. Lu T, Fu C, Ma MH, Fang C, Turner AM. Healthcare applications of smart watches. A systematic review. Appl Clin Inform 2016 Sep 14;7(3):850-869 [FREE Full text] [doi: 10.4338/ACI-2016-03-R-0042] [Medline: 27623763]

12. Srygley JM, Herman T, Giladi N, Hausdorff JM. Self-report of missteps in older adults: a valid proxy of fall risk? Arch Phys Med Rehabil 2009 May;90(5):786-792 [FREE Full text] [doi: 10.1016/j.apmr.2008.11.007] [Medline: 19406298]

13. Casilari E, Luque R, Morón MJ. Analysis of Android device-based solutions for fall detection. Sensors (Basel) 2015 Jul 23;15(8):17827-17894 [FREE Full text] [doi: 10.3390/s150817827] [Medline: 26213928]

14. Lapierre N, Neubauer N, Miguel-Cruz A, Rios Rincon A, Liu L, Rousseau J. The state of knowledge on technologies and their use for fall detection: a scoping review. Int J Med Inform 2018 Mar;111:58-71. [doi: 10.1016/j.ijmedinf.2017.12.015] [Medline: 29425635]

15. Pannurat N, Thiemjarus S, Nantajeewarawat E. Automatic fall monitoring: a review. Sensors (Basel) 2014 Jul 18;14(7):12900-12936 [FREE Full text] [doi: 10.3390/s140712900] [Medline: 25046016]

16. Casilari E, Oviedo-Jiménez MA. Automatic fall detection system based on the combined use of a smartphone and a smartwatch. PLoS One 2015;10(11):e0140929 [FREE Full text] [doi: 10.1371/journal.pone.0140929] [Medline: 26560737]

17. Maglogiannis I, Ioannou C, Spyroglou G, Tsanakas P. Fall detection using commodity smart watch and smart phone. In: Artificial Intelligence Applications and Innovations 10th IFIP WG 12.5 International Conference, AIAI 2014, Rhodes, Greece, September 19-21, 2014. Proceedings. Berlin, Heidelberg: Springer; 2014:70-78.

18. Mauldin TR, Canby ME, Metsis V, Ngu AHH, Rivera CC. SmartFall: a smartwatch-based fall detection system using deep learning. Sensors (Basel) 2018 Oct 09;18(10):3363 [FREE Full text] [doi: 10.3390/s18103363] [Medline: 30304768]

19. Sukreep S, Elgazzar K, Chu H, Mongkolnam P, Nukoolkit C. iWatch: a fall and activity recognition system using smart devices. Int J Comp Commun Eng 2019;8(1):18-31. [doi: 10.17706/ijcce.2019.8.1.18-31]

20. Medrano C, Igual R, Plaza I, Castro M. Detecting falls as novelties in acceleration patterns acquired with smartphones. PLoS One 2014;9(4):e94811 [FREE Full text] [doi: 10.1371/journal.pone.0094811] [Medline: 24736626]

21. Fudickar S, Lindemann A, Schnor B. Threshold-based fall detection on smart phones. In: Proceedings of the International Conference on Health Informatics. 2014 Presented at: HEALTHINF; March 3-6, 2014; Loire Valley, France p. 303-309. [doi: 10.5220/0004795803030309]

22. Silva M, Teixeira PM, Abrantes F, Sousa F. Design and evaluation of a fall detection algorithm on mobile phone platform. In: Gabrielli S, Elias D, Kahol K, editors. Ambient Media and Systems Second International ICST Conference, AMBI-SYS 2011, Porto, Portugal, March 24-25, 2011, Revised Selected Papers. Berlin, Heidelberg: Springer; 2011:28-35.

23. Sposaro F, Tyson G. iFall: an Android application for fall monitoring and response. Annu Int Conf IEEE Eng Med Biol Soc 2009;2009:6119-6122. [doi: 10.1109/IEMBS.2009.5334912] [Medline: 19965264]

24. Awais M, Palmerini L, Bourke AK, Ihlen EAF, Helbostad JL, Chiari L. Performance evaluation of state of the art systems for physical activity classification of older subjects using Inertial Sensors in a Real Life Scenario: a benchmark study. Sensors (Basel) 2016 Dec 11;16(12):2105 [FREE Full text] [doi: 10.3390/s16122105] [Medline: 27973434]

Abbreviations

NB: Naive bayes
SVM: support vector machine

XSL•FO

RenderX

Original Paper

Exploring Patient and Staff Experiences With Video Consultations During COVID-19 in an English Outpatient Care Setting: Secondary Data Analysis of Routinely Collected Feedback Data

Hannah Bradwell[1], PhD; Rebecca Baines[1], PhD; Katie J Edwards[1], MSc; Sebastian Stevens[1], PhD; Kate Atkinson[2]; Ellen Wilkinson[2], MD; Arunangsu Chatterjee[1], PhD; Ray B Jones[1], PhD

[1]Centre for Health Technology, University of Plymouth, Plymouth, United Kingdom

[2]Cornwall Partnership National Health Service Foundation Trust, Cornwall, United Kingdom

Corresponding Author:
Hannah Bradwell, PhD
Centre for Health Technology
University of Plymouth
Drake Circus
Plymouth, PL4 8AA
United Kingdom
Phone: 44 07975927341
Email: hannah.bradwell@plymouth.ac.uk

Abstract

Background: Video consultations (VCs) were rapidly implemented in response to COVID-19 despite modest progress before.

Objective: We aim to explore staff and patient experiences with VCs implemented during COVID-19 and use feedback insights to support quality improvement and service development.

Methods: Secondary data analysis was conducted on 955 patient and 521 staff responses (from 4234 consultations; 955/4234, 22.6% and 521/4234, 12.3%, respectively) routinely collected following a VC between June and July 2020 in a rural, older adult, and outpatient care setting at a National Health Service Trust. Responses were summarized using descriptive statistics and inductive thematic analysis and presented to Trust stakeholders.

Results: Most patients (890/955, 93.2%) reported having *good* (210/955, 22%) or *very good* (680/955, 71.2%) experience with VCs and felt listened to and understood (904/955, 94.7%). Most patients accessed their VC alone (806/955, 84.4%) except for those aged ≥71 years (23/58, 40%), with ease of joining VCs negatively associated with age (*P*<.001). Despite more difficulties joining, older adults were most likely to be satisfied with the technology (46/58, 79%). Patients and staff generally felt that patients' needs had been met (860/955, 90.1% and 453/521, 86.9%, respectively), although staff appeared to overestimate patient dissatisfaction with VC outcomes (*P*=.02). Patients (848/955, 88.8%) and staff (419/521, 80.5%) felt able to communicate everything they wanted, although patients were significantly more positive than staff (*P*<.001). Patient satisfaction with communication was positively associated with technical performance satisfaction (*P*<.001). Most staff members (466/521, 89.4%) reported positive (185/521, 35.5%) or very positive (281/521, 53.9%) experiences with joining and managing VCs. Staff reported reductions in carbon footprint (380/521, 72.9%) and time (373/521, 71.6%). Most patients (880/955, 92.1%) would choose VCs again. We identified three themes in responses: barriers, including technological difficulties, patient information, and suitability concerns; potential benefits, including reduced stress, enhanced accessibility, cost, and time savings; and suggested improvements, including trial calls, turning music off, photo uploads, expanding written character limit, supporting other internet browsers, and shared online screens. This routine feedback, including evidence to suggest that patients were more satisfied than clinicians had anticipated, was presented to relevant Trust stakeholders, allowing for improved processes and supporting the development of a business case to inform the Trust decision on continuing VCs beyond COVID-19 restrictions.

Conclusions: The findings highlight the importance of regularly reviewing and responding to routine feedback following digital service implementation. The feedback helped the Trust improve the VC service, challenge clinician-held assumptions about patient experience, and inform future use of VCs. It has focused improvement efforts on patient information; technological improvements such as blurred backgrounds and interactive whiteboards; and responding to the needs of patients with dementia, communication or cognitive impairment, or lack of appropriate technology. These findings have implications for other health care providers.

Introduction

COVID-19 is a global health concern [1] that has resulted in the rapid implementation and digitalization of many health care services [2,3]. As a result, video consultations (VCs), also referred to as remote or virtual consultations [4], now form an integral part of both primary and outpatient (ambulatory) care. Although limiting viral exposure and reducing the potential risk of infection for patients and staff [4,5], VCs may also enable additional visual cues beyond the capabilities of telephone consultations, helping further facilitate therapeutic relationships and experiences of care [2].

Although there is guidance on how to deliver VCs [6-8] and growing evidence exploring the rapid implementation of VCs in various areas of health care [4,9-17], there are relatively few empirical studies on VCs. For example, Doraiswamy et al [16] reviewed 543 articles related to telehealth (including telephone, VC, and other communication methods) during COVID-19, and only 12% of articles presented empirical work, with few studies conducted in the United Kingdom focusing on VC [16]. Similarly, other research has focused on a single service, such as orthopedics or mental health [4,13,14]. Although the research by Dhahri et al [15] focuses on feedback across a range of outpatient services from 45 patients and 79 clinicians, this was only over a 3-week period.

The emerging research on remote health care and VC implementation seems to show some benefits. For example, VCs are seen as useful for social distancing [12], may provide quicker consultation times [4], reduce travel time for patients [14], and allow for safeguarding and risk assessment [13]. However, the research to date also shows areas of concern, such as technology limitations (impairing video and sound) [4,9,15], additional burden, lack of physical examination [9], low technology confidence and limited setup support [4], and impaired therapeutic interactions and reduced depth of clinical encounters [14].

However, the sample size of VC respondents in studies to date is limited, and no studies to our knowledge have been reported from a British rural and deprived region with an older population [18]. Documented implications of routine feedback on practice have also been lacking in previous literature. Thus, our study aims to investigate National Health Service (NHS) staff and patient experiences with the Attend Anywhere VC in Cornwall using routine feedback from a large sample and to explore the impact of insights shared when presented to key stakeholders in the service.

Methods

Design

This study consists of a secondary analysis of routinely collected, anonymized survey data following a VC that was designed and distributed by the partner health care provider (NHS Trust) and subsequent follow-up to disseminate results and assess the impact of feedback with the Trust.

Setting

The Cornwall Foundation Partnership NHS Trust provides mental health and community services for Cornwall and the Isle of Scilly, a geographically isolated peninsula in the South West of England that experiences higher than average levels of deprivation [19]. By UK definitions, Cornwall and the Isle of Scilly are very rural, with 40% of the population living in remote areas [19], an older age profile [20], and a single acute hospital located in the center of the county. The Trust's services include children and adolescent mental health, adult mental health, and physical health, including but not limited to learning disabilities, cardiac services, bladder and bowel, complex care and dementia, eating disorders, personality disorders, psychiatric liaison, palliative care, stroke nursing, speech and language therapy, diabetes, epilepsy, minor injuries, musculoskeletal care, neurorehabilitation, physiotherapy, podiatry, and respiratory nursing. This study was carried out by an independent research team using anonymous data provided by the Trust.

VC Service

The Trust started using Attend Anywhere, a telemedicine platform for outpatient care [21,22], on April 6, 2020. The platform facilitates video calls between clinicians and patients for scheduled appointment times. The video calls can be conducted over internet-connected computers, phones, or tablets. The implementation process followed guidance provided by NHS England and NHS Improvement [23]. Some appointments were still carried out face to face, but many patients were offered VC or telephone contact during the study period.

Data Collection

The Trust set up a system of routine feedback using a web-based survey (Multimedia Appendix 1) based on the standardized feedback survey questions provided in the national guidance [23]. Immediately after participating in a VC, all staff and patients were invited to complete the web-based survey regardless of whether they had completed one before. The survey was presented via the Meridian surveying platform at the end of a VC. To our knowledge, the survey was not designed with patient or public involvement owing to the rapid implementation process. The rapid rollout of this feedback process may also be responsible for some limitations in the surveys themselves, including limitations in response options and understanding of the sample (eg, number of unique

individuals and complete understanding of the service accessed by the patient). However, we have confidence that the data items used in the analysis are robust based on their recommended use by NHS England and NHS Improvement.

Participants

Participation was voluntary, and the patient participants were patients or their carers. Survey respondents gave consent for their data to be used, but some chose not to have comments publicly shared. Their data were included in the analysis and production of the themes, but their quotes were not included. We used data collected during the early implementation of VCs in response to COVID-19 (June 1, 2020, to July 31, 2020). During this time, 4234 Attend Anywhere appointments were completed. The feedback data used were 22.56% (955/4234) patient and 12.3% (521/4234) staff responses out of the 4234 completed VCs. The sample size was largely pragmatic, using data from as many patients (nearly 1000) as was thought possible to thematically analyze in a timely manner to give feedback to staff during a 2-month period after an initial *settling down* of the system but early enough to have practical use in assessing the utility of the method.

Data Analysis

Descriptive statistics were reported for numerical data, and chi-square and Mann–Whitney *U* tests were used where appropriate. Free-text responses were analyzed by 4 researchers (HB, RB, KE, and SS) using inductive thematic analysis [24]. Initially, staff and patient comments were analyzed separately. A comprehensive coding framework was developed. Patient and staff codes demonstrated high comparability and, thus, are presented as combined themes across the data set, with areas of discordance discussed. Thematic analysis was selected as a useful and flexible method to generate a rich yet detailed and complex account of qualitative data [24]. Adopting an inductive approach also helped ensure that identified themes arose from the data generated as opposed to predefined concepts or ideas.

Documented Impact of Routine Feedback on Real-world Practice

Following analysis of the results, summary presentations were created and presented to relevant stakeholders as rapid feedback between November 24, 2020, and February 21, 2021. Presentations were given in partnership with a service-user consultant to patients, patient representatives, and professionals at an Experiences of Care collaborative meeting within the care system, the South West Outpatient Transformation group, the

region's VC forum, and local web-based research dissemination events, and to national audiences through the Outpatient Transformation regional leads meeting. The results were also shared with interested international health care providers (Finland).

Patient and Public Involvement

The research question and study were informed by patient input (through inclusion of patient experience), and a service-user consultant contributed to reviewing and coauthoring the manuscript.

Ethics Approval

Ethical approval to conduct this secondary analysis was provided by the Health Research Authority and Health and Care Research Wales, Integrated Research Application System ID286543 (27.07.2020). This manuscript was prepared using the Standards for Quality Improvement Reporting Excellence guidelines [20].

Data Sharing Statement

Not all patients consented for their comments to be published; thus, the full data set is not publicly available. However, interested parties may inquire with the authors for further details.

Results

The results are presented in three sections: (1) quantitative results, including participant characteristics; (2) qualitative results; and (3) documented impact of this routine feedback on real-world practice. Multimedia Appendix 1 provides further details on the questions presented to staff and patients in the survey.

Quantitative Results

Patient Age and Device Used

Just under a quarter (955/4234, 22.56%) of the 4234 patient VCs resulted in feedback responses. As the data were anonymous, it was not possible to know if some individuals completed the survey more than once in this 2-month period. Therefore, each data entry was treated as an individual episode. The highest number of survey responses was received from individuals aged 31-50 years (333/955, 34.9%), with the lowest response from patients aged >71 years (58/955, 6.1%). Half of the patients (487/955, 51%) used a laptop to access their VC. Devices used varied by age ($\chi^2_{12}=68.9$; *P*<.001)—patients aged >50 years were much less likely to use mobile devices and more likely to use a tablet (Table 1).

Table 1. Use of devices for the video consultation by age group, showing numbers (percentages; N=955).

Item	Device, n (%)				Total, n (%)
	Laptop	Mobile phone	Tablet	Other	
Age group (years)					
<18	82 (50.3)	44 (27)	32 (19.6)	5 (3.1)	163 (17.1)
18-30	76 (52.1)	35 (24)	24 (16.4)	11 (7.5)	146 (15.3)
31-50	167 (50.2)	87 (26.1)	64 (19.2)	15 (4.5)	333 (34.9)
51-71	120 (49.6)	20 (8.3)	70 (28.9)	32 (13.2)	242 (25.3)
>71	36 (62.1)	1 (1.7)	16 (27.6)	5 (8.6)	58 (6.1)
Total	487 (51)	189 (19.8)	209 (21.9)	70 (7.3)	942[a] (98.6)

[a]A total of 13 respondents did not answer regarding age group.

Staff Characteristics

In total, 521 staff responses were received (Table 2), with a response rate of 12.3% (521/4234). The largest number of responses by profession was from Allied Health Professionals (155/521, 29.8%). The largest number of responses by department was from Community Mental Health (188/521, 36.1%). Most staff responses were completed following VCs with a patient for whom staff members had 1-3 previous contacts. Staff data, such as patient data, are episodes rather than individuals.

Table 2. Staff respondents, showing profession and department (N=521).

Characteristics	Participants, n (%)
Profession	
Nurse	81 (15.5)
Psychologist	117 (22.5)
AHP[a]	155 (29.8)
Physician	61 (11.7)
Other	107 (20.5)
Department	
ACS[b] inpatient	17 (3.3)
ACS community	114 (21.9)
Community Mental Health	188 (36.1)
Mental health inpatient	4 (0.8)
CAMHS[c]	127 (24.4)
Children's Services	56 (10.7)
Complex care and dementia	10 (1.9)
AMH[d] and learning disabilities	5 (1)

[a]AHP: Allied Health Professionals.

[b]ACS: Adult Community Services (eg, podiatry, spinal, physical, and rehabilitation).

[c]CAMHS: Child and Adolescent Mental Health Services.

[d]AMH: Adult Mental Health.

Patient Overall Experience and Future Intention

Most patients (890/955, 93.2%) reported having a *good* (210/955, 22%) or *very good* (680/955, 71.2%) overall experience with VC. A small number of patients had a *poor* (13/955, 1.4%) or *very poor* (17/955, 1.8%) experience. Future intention could also be seen as a measure of satisfaction with VC—9 out of 10 patients were *very likely* (704/955, 73.7%) or *somewhat likely* (176/955, 18.4%) to choose a VC in the future. Very few patients (28/955, 2.9%) suggested they were *somewhat unlikely* (17/955, 1.8%) or *very unlikely* (11/955, 1.2%) to use VCs in the future. Within the results, we were able to look at two aspects of overall satisfaction: satisfaction with the technology (video and sound) and satisfaction with the communication (more related to the clinician's performance in this situation).

XSL•FO

RenderX

Patient and Staff Technical Satisfaction

Three-quarters of the patients reported having a *very positive* experience with sound and video quality (732/955, 76.6% and 728/955, 76.2%, respectively). When combined, 67.6% (646/955) had a very positive experience with both video and sound (Table 3). Three-quarters of staff also reported a *positive* or *very positive* experience with sound and video quality (411/521, 78.9% and 399/521, 76.5%, respectively).

Table 3. The 4 indicators of patient satisfaction with video consultations shown by patient age and device used, showing *P* values from chi-square test (N=955).

Characteristic	Positive overall	Would choose VC[a] again	Positive about technology	Positive about communication	Total patients
Age (years)					
<18, n (%)	153 (93.9)	146 (89.6)	94 (57.7)	137 (84)	163 (17.3)
18-30, n (%)	133 (91.1)	134 (91.8)	92 (63)	131 (89.7)	146 (15.5)
31-50, n (%)	316 (94.9)	313 (94)	240 (72.1)	290 (87.1)	333 (35.4)
51-71, n (%)	224 (92.6)	225 (93)	168 (69.4)	200 (82.6)	242 (25.7)
>71, n (%)	55 (94.8)	53 (91.4)	46 (79.3)	53 (91.4)	58 (6.2)
Total, n (%)	881 (93.5)	871 (92.5)	640 (67.9)	811 (86.1)	942[b] (100)
P value	.55	.49	.003[c]	.18	N/A[d]
Chi-square (*df*)	3.0 (4)	3.4 (4)	15.8 (4)	6.2 (4)	N/A
Device					
Laptop, n (%)	445 (91.4)	441 (90.6)	307 (63)	407 (83.6)	487 (51)
Mobile phone, n (%)	184 (97.4)	179 (94.7)	147 (77.8)	168 (88.9)	189 (19.8)
Tablet, n (%)	200 (95.7)	195 (93.3)	152 (72.7)	183 (87.6)	209 (21.9)
Other, n (%)	61 (87.1)	65 (92.9)	40 (57.1)	60 (85.7)	70 (7.3)
Total, n (%)	890 (93.2)	880 (92.1)	646 (67.6)	818 (85.7)	955 (100)
P value	.003[c]	.28	<.001[c]	.27	N/A
Chi-square (*df*)	13.8 (3)	3.9 (3)	19.6 (3)	3.9 (3)	N/A

[a]VC: video consultation.

[b]A total of 13 missing ages.

[c]*P*<.05.

[d]N/A: not available.

Patient and Staff Satisfaction With Communication

Most patients felt that they had been listened to and understood (904/955, 94.7%), had had their needs met (860/955, 90.1%), and had been able to communicate everything they wanted (848/955, 88.8%). Overall, 85.7% (818/955) of the patients rated all 3 aspects positively (Table 3).

Most staff members (419/521, 80.4%) felt able to communicate everything they wanted to, although satisfaction was slightly lower than that of patients (419/521, 80.4% vs 848/955, 88.8%; χ^2_1=19.4; *P*<.001). Staff perceptions of patients feeling their needs were met were generally positive, with 57% (297/521) and 29.8% (155/521) responding *yes* and *yes partially*, respectively. Only 2% (19/955) of patients said their needs were not met, whereas 11.1% (58/521) of staff believed that patients felt their needs were not met, suggesting an apparent discrepancy.

Association Between Patient Satisfaction With Technology and Communication (Combined as Above)

Patient satisfaction with communication was very strongly positively associated with satisfaction with technical performance (χ^2_1=104.0; *P*<.001; Table 4). Only 4.3% (41/955) of patients were less satisfied with the communication despite being satisfied with the technology. Conversely, 22.3% (213/955) of patients remained positive about the communication despite being less positive about the technology.

Table 4. Cross-tabulation of patients being positive about technology with patients being positive about communication (N=955).

Positive about technology	Positive about communication, n (%)		Total, n (%)
	No	Yes	
No	96 (70.1)	213 (26)	309 (32.4)
Yes	41 (29.9)	605 (74)	646 (67.6)
Total	137 (100)	818 (100)	955 (100)

Patient Independence and Accessibility

Most patients (806/955, 84.4%) stated that they could access their VC alone and that it was *very easy* or *easy* to join the VC call (849/955, 88.9%). However, more respondents aged >71 years reported needing help (35/58, 60%), and fewer reported it easy to join compared with those aged <31 years (40/58, 69% vs 286/309, 92.6%; χ^2_3>=28.6; P<.001).

Influence of Patient Age and Device Used on Outcome

The relationship between device used, age, and satisfaction with technology and the VC was complex. On the one hand, older patients were more positive about the technical experience despite being more likely to need help accessing the VC and being less likely to find accessing the VC easy. Mobile users were also more positive (Table 3), and older adults were less likely to be mobile users (Table 1).

Staff Experience With Managing and Joining the Call

Most staff respondents reported a *positive* or *very positive* experience when managing and joining the VC (466/521, 89.4%). A smaller number of staff responses indicated a *negative* or *very negative* experience (49/521, 9.4%).

Patient and Staff Perceived Savings

Two-thirds of patients reported a perceived saving in time (662/955, 69.3%), with more than half of respondents also reporting a perceived saving in money (544/955, 57%). There was no difference by age. Staff respondents most commonly identified carbon savings (380/521, 72.9%) followed by time (373/521, 71.6%). Over one-third of staff reported saving money (187/521, 35.9%). Just below one-quarter of staff respondents reported a perceived saving on missed appointments or *did not attends* (DNAs; 128/521, 24.6%). Approximately 24% (125/521) reported *other* unspecified savings. Savings are explored in greater depth in the qualitative analysis below.

Patient Versus Staff Perception

Overall, there appears to be good concordance between staff and patient feedback, with similar benefits noted for time and money savings. Mann–Whitney *U* tests demonstrated no significant difference between 521 staff and 955 patient ratings of video (P=.15) or sound quality (P=.77). However, significant differences between staff and patient responses were identified when reviewing whether patients had been able to communicate everything needed and felt their needs were met. On both occasions, staff responded more negatively than patients (P<.001; P=.02). This could suggest that staff overestimated patient dissatisfaction with VC outcomes or were not aware of patient experiences. This is a useful finding, which was reported back to stakeholders at the Trust; however, limitations in the survey (as discussed in Multimedia Appendix 1) must also be considered.

Qualitative Results

Number of Comments

Overall, 13.9% (133/955) of patients made 1384 free-text comments in response to one or more of the 16 questions (Multimedia Appendix 1). Patients who rated their overall experience as good or very good were much less likely to comment (105/890, 11.8% vs 28/65, 43.1%; P<.001). Two-thirds (350/521, 67.2%) of staff made 528 free-text comments in response to one or more of the 9 questions asked.

Overall Themes

Overview

Inductive thematic analysis of free-text responses identified three main themes: barriers, benefits, and suggested improvements. Although the overarching themes were the same based on staff and patient analysis, there was some variation in initial codes between the 2 groups, as discussed in the narrative below and shown in Table 5. Unique identifiers are used for each quote.

Table 5. Qualitative patient feedback themes, subthemes, and codes.

Theme and subtheme	Codes
Barriers	
Technological issues	• Equipment, sound or video issues, difficult to communicate, connectivity issues, sound quality, video quality, joining issues (SP[a]) • Impaired therapeutic flow, limited support, poorer-quality interactions, increased staff stress (S[b])
Quality of patient information and administrative support	• Jargon, accuracy, complexity of language (P[c]) • Lack of technical support, human error, patient struggles with VCs[d] and joining (SP)

Accessibility and suitability concerns	• Lack of suitable or compatible devices and up-to-date browsers, support required (SP)
	• Widening inequalities, difficult to risk-assess, no hands-on care, suitability of certain conditions including the following (S):
	  • Hearing-impaired (SP)
	  • Children and attentional issues (S)

Time, resource, and cost concerns	• Increased personal cost for staff, increased staff time, more DNAs[e] (S)
Benefits	
Reduced anxiety and stress	• Comfort, face-to-face element, relaxing and anxiety-reducing, safer, patients better supported, family members present (SP) • Ability to open up more (S)
Continued service delivery	• Allows examination, facilitates contact with patients and staff, higher-quality appointments, nonverbal cues, safety, continued service (S)
Perceived savings	• Travel, money, time, environment, work hours, arranging lifts, childcare (SP)
Enhanced accessibility	• Increased access, affordability of attending, comfort, childcare, arranging lifts (SP) • Fatigue (P)
Suggested improvements	
Information and support improvements	• Notify if appointments are running late (P) • Allow trial VC, provide reminders, simplify or improve patient information, simplify sign-in process, device advice (SP) • Camera positioning guidance (S)
Technological improvements	• Turn off music, allow photo upload before appointment, support use in other browsers, expand character limit, allow document editing (P) • Shared interactive whiteboard, resources or activities, background blur for privacy, allow control by patient (S)

[a]Codes present in staff and patient data.

[b]Codes that resulted only from staff data.

[c]Codes that resulted only from patient data.

[d]VC: video consultation.

[e]DNA: did not attend.

Barriers

The participants related barriers from technological difficulties, quality of patient information, and concerns about accessibility or suitability of using VC.

Technology

Many patients identified concerns of connectivity—"platform glitches" [Patient 94]—or experienced delays between video and sound:

> *The picture kept freezing and pixelating.* [Patient 741]

> *The video quality is so poor it's really hard to get much done.* [Patient 332]

In some cases, technical difficulties meant patients felt that their "needs were not met" [Patient 305]. Similarly, some clinicians reported having to "resort to a telephone" [Staff 14] to supplement VC audio. Some staff members reported that VCs had "a very negative effect on the quality of...therapy we can deliver" [Staff 28] and "limit...the complexity of the conversation" [Staff 29], making it "hard to pick up on body language" [Staff 144]. For some staff members, impairments

caused by technological issues were seen as "detrimental to patient care" [Staff 90], "frustrating" [Staff 197, Staff 409, and Staff 439], and "stressful" [Staff 90]. It is possible that the connectivity issues seen in this study related to the geographical character of the region of Cornwall, with staff reporting the following:

> *The general Cornish bandwidth is the obstruction here.* [Staff 118]

However, staff reported other platforms seemed to encounter fewer issues, suggesting the platform "needs to be improved substantially and quickly" [Staff 22].

Quality of Patient Information and Administrative Support

The quality of patient information, particularly joining instructions, was repeatedly called into question by both patients and staff. A patient described the joining process as "stupidly complex" [Patient 52]. Others described being "directed to a troubleshooting, jargon filled, suggestions page" [Patient 95]. The accuracy of the joining instructions was also questioned:

> *The link doesn't open as the instructions [suggest].* [Patient 236]

Although a video version of the information was provided, this was described by the participants as in need of further development and refinement to ensure inclusivity, particularly for "deaf patients" [Patient 582]. Similarly, staff felt that the "main issues" [Staff 75] for patients involved "logging on" [Staff 71] and that "the process of joining has lots of information to process" [Staff 75]. The consequence being "patients [take] a while to get into the appointment [as] the process [is] complicated" [Staff 150]. Some clinicians needed to "telephone and talk [the patient] through logging on" [Staff 286].

Accessibility

Access to relevant devices, browsers, digital skills, and confidence were also described as problematic by some participants:

> *Unfortunately I was not up to date with technology.* [Patient 319]

Some patients reported needing to download alternative browsers or borrow other people's devices because of "outdated" [Patient 671] models. Staff also suggested the following:

> *It can be hard to engage those with limited IT equipment.* [Staff 106]

Similar to the quantitative findings outlined above, some participants reported needing help from family members or friends as "without [them] it would have been impossible" [Patient 540]. Another "96 year old had to pay for a carer to be present [and] 2 hours of IT help from someone else" [Patient 47], raising further questions and concerns.

Suitability Concerns

Although patient concerns focused mainly on technology and digital exclusion, staff had additional concerns about suitability based on patient illness or requirement. Some staff members felt that VCs were exacerbating "health inequalities" for individuals with learning disabilities or living in residential homes as patients were "often excluded from the review" [Staff 202] as computers were often located "in an office" [Staff 202]. This concern was echoed, as "clients with learning disabilities" [Staff 15] often "need reasonable adjustments to be facilitated to communicate" [Staff 15]. The usefulness of VCs for dementia services was also queried, with the "screen [removing] sensory aspects and visual clues" [Staff 130]. VCs were also seen as unsuitable for dysphagia, where "a hands on approach is required to closely look, listen and feel as the person eats and drinks" [Staff 25]. Patient conditions that may impede VC success as suggested by staff respondents also included "cognitive, speech, language, fatigue, concentration, need for physical...assessments, environmental assessments, safety" [Staff 23]. Other areas described as problematic included family therapy, where the "family had to sit side by side, so parents couldn't see their child's facial expressions" [Staff 278], and VCs with children generally, particularly with attentional needs, which "meant [the] session was longer" [Staff 136] and fewer tasks were achieved than face to face. Patient safeguarding and "environmental assessments" [Staff 23] also appeared to be a key issue for clinicians—"home visits remain hugely important to gather information to ensure patient safety," [Staff 130] providing more information "such as how a person may be living and identify self-neglect, declines in functional skills or poor medication management" [Staff 130]. Staff also suggested some conversations may "be very challenging" [Staff 202] over VC, such as discussing "the risk of dying over the internet and not in person" [Staff 202], as some respondents felt "discussing end of life care" [Staff 29] over video carried "an increased risk of missing cues" [Staff 29]. Patients raised concerns for people with hearing impairments as "face to face is easier" [Patient 457] for lip readers because of video or sound delays and character limitations in the VC chat function.

Time, Resource, and Cost Ineffectiveness

A minority of staff members reported concerns of time, resource, and cost ineffectiveness as, although VCs "saved a few minutes walking to the clinic and tidying the room for the assessment," [Staff 169] they could also "add an extra appointment" [Staff 169] when a face-to-face consultation was needed. Some respondents also suggested that VCs were longer owing to a "longer explanation time" [Staff 279] to talk patients through the process of how to log on, as described above. Technological issues also affected duration, with patients "having to change rooms" [Staff 198], "change to telephone" [Staff 203], or "re-join" [Staff 19] after disconnection. Furthermore, staff felt VCs "take quite a bit of preparation prior to the consultation" [Staff 23], particularly in psychological services where staff may "need to make electronic versions of therapy resources" [Staff 6]. Although 24.6% (128/521) of staff reported reduced DNAs numerically, some staff members suggested DNAs had increased:

> *I've had more DNAs than when most of my visits were by car.* [Staff 413]

However, only 3 such comments were made. Interestingly, patients did not report time or resource ineffectiveness and generally reported savings, as described below.

Interestingly, none of the patients described the therapeutic relationship or quality of care delivered by individual health care professionals as a barrier or limitation of VC use.

Benefits

Overview

The participants identified benefits of continued high-quality service delivery, including reduced anxiety and stress, perceived savings, and enhanced accessibility. Most patients repeatedly described their VC experience as "fantastic" [Patient 558], "excellent" [Patient 579], "amazing" [Patient 153], "wonderful" [Patient 863], "positive" [Patient 600], and "useful" [Patient 667]. The participants also appeared to appreciate being able to have "family members join the appointment" [Patient 15].

Continued High-Quality Service Delivery

For staff, VCs allowed them to "continue to provide care" [Staff 47] and "maintain contact and show patients that we are here, that we are holding them in mind and we are motivated to help" [Staff 24]. Clinicians noted that, without VCs, provision would be "even more reduced" [Staff 47] and that services such as "psychotherapy" [Staff 160] did not need to go "on hold" [Staff 160]. Staff who were "shielding" [Staff 356] also managed to maintain workloads and provide care for "shielding patients" [Staff 394]. Some clinicians reported good "depth of therapy work," which was "less tiring than therapy by telephone" [Staff 60].

Reduced Anxiety and Stress

Both patients and staff felt that VCs were "far less stressful" [Patient 796] than face-to-face consultations. Several patients reported feeling "more relaxed" [Patient 563] and less "rushed" [Patient 392] as a result of avoiding certain stressors, including arranging transport, arriving on time, finding and paying for parking, and traffic. Reduced anxiety and stress were also reported among children, particularly in children who experienced anxiety about appointments or leaving the house. Clinicians noted that child assessments could be supported by the presence of teaching staff in addition to parents alone and noted the following for adults with learning disabilities:

> *Distant parents [were] able to join the review, whereas they wouldn't have previously.* [Staff 202]

Staff also reported that VCs were "less stressful" [Staff 89] for the clinicians themselves, "increasing my wellbeing" [Staff 64] by "saving stress" [Staff 209] and "anxiety and distress" [Staff 519]. Other staff members suggested that VCs "can actually be more therapeutically productive" [Staff 170] with "improved communication" [Staff 110] and "better clinical contact" [Staff 200]. Patients reported that the removal of "so much stress" [Patient 290] meant that they had "more time to focus on what needs talking through" [Patient 290]. Several participants suggested that, as they "didn't feel so stressed" [Patient 695] and were in the comfort of their "own home" [Patient 564], they were "able to open up more" [Patient 695], often feeling more "comfortable" [Patient 564] and "relaxed" [Patient 564]. Similarly, for staff, patients receiving care at home was seen as beneficial, such as for a "post-natal mum" who "felt more comfortable in their own home" [Staff 126]. VCs were also seen as reducing "anxiety for patients concerned about face-to-face

appointments due to COVID" [Staff 29] and allowed for "concordance" [Staff 58] in patient care.

Perceived Savings

Both patients and staff reported time, monetary, and environmental savings. Some patients reported saving "over £20 in transportation costs" [Patient 98] and being able to now "afford" [Patient 153] an appointment as a result of time and cost savings:

> *[VCs] have genuinely changed my life...being accessible for my needs and being able to afford an appointment.* [Patient 153]

Travel and cost savings may be particularly prevalent in "Cornwall," where it "is always difficult to travel for appointments" [Patient 262]. Many staff members also suggested VCs save "time, money and travelling for all concerned" [Staff 324], also reducing "carbon footprint" [Staff 75] by saving "on travel" [Staff 315], "paper" [Staff 242], and "printing resources" [Staff 261]. Many clinicians reported that they could "see patients more intensively" [Staff 7] and "complete an increased number of appointments in a day" [Staff 19] with "increased capacity as a whole" [Staff 17]. Clients who were often "late due to travel" [Staff 81] were now on time. Clinicians reported saving "90 minutes in the car visiting a patient who is just as happy to be seen by video" [Staff 307]. Some staff members also reported that they "rarely get DNAs" [Staff 444] and that VCs "must have saved [The Trust] a lot of money" [Staff 444].

Enhanced Accessibility

Although digital exclusion was thought to reduce accessibility for some, as outlined above, most free-text responses suggested that VCs facilitated service accessibility in a number of ways. First, owing to certain conditions and reduced mobility, some respondents found "trips to the hospital very tiring and difficult" [Patient 20]. VCs removed this experience. Childcare and employment cost savings were also described, helping increase accessibility. For example, a participant suggested that, for a face-to-face appointment, they "would have been dragging all three kids along for what ended up being something that the video call was able to address" [Patient 863]. The "option of video call" [Patient 335] was also considered "useful" [Patient 335] for those who are employed or "working parents" [Patient 335]. Patients could schedule weekly appointments around their employment, something that is not always possible when relying on face-to-face appointments. Furthermore, a number of patients who were "not able to drive" [Patient 307] or "can't drive" [Patient 846] described VCs as "much more convenient" [Patient 307] because of enhanced independence and removal of reliance on others. Similarly, staff reported increased accessibility for numerous patients who ordinarily "cannot travel to appointments" [Staff 77], such as those with "mobility issues" [Staff 201]. The lesser time requirement for patients to attend via VC was considered to increase "availability of services to clients with other commitments" [Staff 274], "work engagements" [Staff 153], or "caring responsibilities" [Staff 280].

As a result of the benefits encountered, many patients expressed a strong desire for VCs to be made available beyond the COVID-19 pandemic:

> I hope that all of our future appointments will be held this way. [Patient 863]

This highlights an important element of patient choice.

Suggested Improvements

Finally, staff and patients suggested improvements in two main areas: information and technology.

Information Improvements

Respondent suggestions included the simplification of patient guidance and information, device-specific advice, and suggested device use for optimum VC experiences. For example, some patients reported positioning their "iPad onto the floor, so I could show...my feet and me walking" [Patient 248]. This would have been less feasible if using "a desktop computer" [Patient 248]. Other participants suggested it was "a bit challenging getting [the] camera in position to demonstrate me doing the exercises" [Patient 809]. Therefore, providing device-specific information and recommending particular devices based on service requirements and availability may be beneficial. Patients also requested some method of notification if the clinics were running late.

Technology Improvements

Related to platform functionality, patients requested the ability to "turn off the music while sitting in the waiting room" [Patient 388] as it was considered "terrible" [Patient 517] and repetitive by some. Patients also expressed a desire to be able to "upload photos or videos" [Patient 20] to the consultation and "change the mobile camera being used" [Patient 145]. Further platform-related improvements suggested included expanding its functionality to "other browsers" [Patient 342] and having the character limit expanded beyond 200 characters for patients who need to use the text function. Other suggestions made by staff included more interactive screen-sharing capabilities as currently clinicians "cannot see [the] client anymore" [Staff 6] when sharing their screen. Others thought it would be useful if the clinician could "allow the client to take control of a particular programme you are sharing," [Staff 6] which could help with "engaging children so you could play games together" [Staff 6]. A similar desire was noted for adult cognitive behavioral therapy, with staff requesting a "white board" [Staff 119] to "draw things out on such as CBT formulations" [Staff 52]. The absence of such functionalities meant that diagrams were completed less "collaboratively" [Staff 107] than if patients were "in the room" [Staff 107]. Finally, related to digital skills and confidence, some patients expressed a desire for a "dummy run" [Patient 74] to be made available so that people could familiarize themselves with the technology before their consultation. Interestingly, no suggestions for health care training were proposed by the participants, although this may reflect the questions asked in the feedback survey. For enhanced security, staff requested a "blurred background" [Staff 3] option "to help protect privacy" [Staff 3] and "improve confidentiality" [Staff 375]. This seemed particularly relevant for clinicians who "work with forensic clients" [Staff 3]. An additional

improvement would be having the ability to "lock the room once everyone is in" [Staff 8] after experiencing "incidents of other staff joining private, confidential therapy sessions uninvited" [Staff 8].

Comparison of Patient and Staff Feedback

Generally, staff and patient responses showed high congruence, as evidenced by the similarity of codes and subsequently combined themes. However, as with the quantitative results, where more staff members responded with concern about patients feeling their needs were met than patients themselves, the qualitative data also demonstrate some evidence of clinicians believing that VCs impaired the therapeutic flow or produced poorer-quality interactions. Patients provided no indication of dissatisfaction with the clinician's communication, outcomes, or care received other than issues resulting from the technology. Although staff concerns on meeting patients' needs also commonly resulted from technical issues, generic concerns were also shared on therapeutic quality and missing cues or body language via VC. In addition, patients regarded highly the increased accessibility of health care and found appointments less stressful. The codes on time, resource, and cost ineffectiveness were provided only by staff, whereas patient data strongly supported perceived savings across time, money, travel, and environmental impact. Staff and patients showed similarity in the reported requirement for improved sound and video quality. In addition, patients requested the removal of waiting room music, expansion of character limits, and trial runs, whereas staff requested blurred backgrounds and interactive shared screens. Both patients and staff requested simplified or improved patient information.

Documented Impact of Routine Feedback on Real-world Practice

Following the initial analysis of the results, the authors presented the above findings to relevant patient and professional stakeholders to provide rapid feedback on barriers and facilitators, possible areas for improvement, and ways to encourage the sustainable use of VCs. The findings were positively received by both clinicians and patients. In particular, clinicians reported underestimating patient satisfaction with VCs and were surprised to see such high levels of satisfaction with the service, particularly patient perceptions of communication quality and feeling their needs were met. The presentations aided in revising clinician perceptions, and some initial changes to VC practice have already been instigated, including the replacement of waiting room music with bird songs and efforts to improve patient information. Thus, the use of routine feedback and its analysis was instrumental in instigating some initial improvements for the use of VCs within the NHS Trust and promoting further conversations around future VC use and improvements. The implementing NHS Trust has now confirmed the procurement of Attend Anywhere for VC service provision to continue beyond COVID-19 restrictions.

Discussion

Principal Findings

This research contributes to the existing literature exploring staff and patient experiences with the rapid rollout of VCs for outpatient care during the COVID-19 pandemic [21], with clear implications for policy, practice, and future research. Specifically, this work contributes to the VC evidence base with a considerably larger sample than previous evaluations, in a rural setting and with a focus on VCs rather than telephone. We also provide documentation on the impact of routine feedback. Unlike existing literature mainly focusing on primary care or specific health services, this research explored the use of VCs across outpatient services more broadly, helping address limitations within the existing COVID-19–response literature [4,10].

Most patients rated their VC experience as *good* or *very good*, felt listened to and understood, were able to communicate everything they wanted to, and felt their needs had been met. Many patients reported saving time and money, and >90% (880/955) would likely choose a VC in the future, although it remains unclear whether this resulted from no perceived *alternative* option owing to the global pandemic or from positive patient experiences and related motivations. Further exploration would be beneficial, including analyzing patient experience over time. Staff also generally supported the use of VCs through positive experiences with joining and managing calls, being able to communicate all that was needed, and feeling the patients' needs had been met, although agreement on the latter issue was not unanimous. Staff also reported savings, mostly in terms of carbon footprint and time. Qualitatively, both staff and patients noted increases in service accessibility and affordability. As a result, this Trust commissioned Attend Anywhere for future (COVID-19 and post–COVID-19) use, and other Trusts should consider making VCs a permanent option beyond pandemic restrictions.

Patients aged ≥71 years (58/955, 6.1%) were the only age group in which a larger percentage of respondents reported needing support accessing their VC than those who were able to access alone. Nevertheless, there was no clear age gradient in satisfaction with VCs. Indeed, older adults were more likely to be positive about the technology than younger people. This may reflect higher expectations among younger people. Although high levels of positive experience were reported across all device types, more users of tablets or mobile devices were positive. The relationship between age, device used, and positive experience with VCs is complex. Research suggests that the small size of mobile phones can pose a barrier for older adults coupled with declining dexterity and vision. Mobile phones are mainly used by older adults for calls and texting [25], whereas tablet computers are more popular for web-based access among older adults than among younger people [26]. Older working age groups may be more likely to prefer technologies they are familiar with at work (often desktops with poor cameras and sound systems), whereas older adults new to computing use tablets as their entry device [27].

Clinical areas of less suitability for VCs were also noted, particularly by staff. Further research is needed to identify when VCs work best, for whom, and in what context. Although Greenhalgh et al [7] have already provided guidance on appropriate and inappropriate use of VCs, a more granular understanding from both a patient and professional perspective may be required. Some limitations were noted for spinal services, neurology, children, attentional issues, and assessing dysphagia. The previous guidance [7] suggested that inappropriate contexts included patients at high risk, patients requiring internal examination, and patients with challenges affecting the ability to use technology. This suggestion is supported by our work, with staff suggesting VC was less appropriate for patients with learning disabilities, communication disorders, fatigue, cognitive issues, or dementia.

Barriers described by staff and patients included technological difficulties, quality of patient information, administrative errors, and accessibility or suitability concerns. Conversely, identified benefits included reduced stress and anxiety for patients and staff, the opportunity to "open up more" for patients as a result of enhanced comfort, cost and time savings, increased sense of affordability, and service accessibility. Finally, the participants suggested a number of improvements, such as simplifying patient information, notifications for late appointments, the ability to turn off waiting room music, a shared interactive whiteboard, blurred backgrounds, and "practice run" opportunities to increase familiarity and digital skill confidence. Other suggested improvements included allowing photo or video upload to the appointment, swapping between cameras used, extending Attend Anywhere to other internet browsers, and expanding the character restrictions of the chat function, particularly important for accessibility of deaf patients. Thus, this study has a number of practical implications. Routinely collecting and responding to feedback is likely to be an integral aspect of service improvement, as demonstrated in this study. Feasible improvements such as those reported here are likely to have important impacts on staff and patient experience.

The need to simplify and improve patient information was highlighted as a key barrier by both staff and patients, and this may be best achieved in co-design with patients. Although the implementation of VCs was a rapid response, actively involving patients and the public and creating digital-related information may improve accessibility, relevance, and understanding. Thus, an implication of this work is an identified need to establish the best practice for rapid co-design when implementation timing is critical. Any future patient information may also include guidance for patients on camera positioning to reduce another barrier identified in this work. Health care services may also benefit from recommending particular devices based on their functionality and service requirements. For example, larger, static screens may be suitable for child therapies or family-based interventions where patients and families need to sit side by side. Alternatively, VCs that include assessment of movement may be better suited to more portable devices such as mobile phones or tablets.

An important consideration for VCs is safeguarding. Bhardwaj et al [13] reported that clinicians were confident in performing safeguarding and risk assessments remotely. Our results indicate

otherwise as clinicians reported that home visits were key for patient safeguarding and to allow monitoring of self-neglect, decline in well-being, or poor medication. Therefore, home visits for patients requiring environmental assessments could be prioritized for face-to-face appointments, as could consultations relating to the identified less appropriate contexts. However, this assumes that environmental assessments are not possible via VC, where perhaps a visual tour of the home environment during a VC would suffice. Thus, an alternative implication is guidance for clinicians in this regard. A further concern was raised by clinicians consulting with patients in residential care with respect to widening health inequalities as patients are often unable to attend consultations where computers are housed in staff offices. In addition, some patients in these contexts appeared unable to appropriately position their camera, suggesting the potential for solutions such as affordable telepresence devices or robotics in residential care to facilitate VCs. This could respond to two barriers: (1) patients being excluded from VCs because of equipment in staff offices and (2) challenges regarding appropriate camera positioning.

As the final practical implication, in response to patients noting administrative and human errors, the collaborative development of checklists and supportive training may be beneficial. Trusts could perhaps include on-screen checklists on patient records to ensure that scheduling of VCs is followed by provision of an appropriate link, patient-facing guidance, and setup support. Clinicians may also consider promoting the benefits of teachers or distant relatives attending a VC.

Our study also raises implications for the collection and use of routine feedback. The clinicians in our sample overestimated patient dissatisfaction with VCs. More clinicians than patients also responded negatively to communication quality. A minority of clinicians reported some impairment of therapeutic flow. Our presentation of these results to Trust stakeholders supported this observation, with clinicians surprised about high patient satisfaction. Some clinicians reported avoiding VCs for fear of patient dissatisfaction. Thus, the provision of this routine feedback aided in addressing staff perceptions. It is possible that low staff expectations for VCs somewhat explains the low documented uptake of VCs in comparison with telephone calls in previous research [10,13,14]. When collecting routine feedback, critically considering the purpose is important. For example, Sibley et al [28] recently likened the increasing collection of patient feedback to an "avalanche...with experience now tracked, monitored and measured to an almost obsessive degree" [28]. However, to what end and for what purpose? Reflecting previously acknowledged concerns around the ethics of collecting patient feedback that leads to minimal direct benefit [29,30], Sheard et al [31] suggested all patient feedback tools must be meaningfully usable by those providing frontline care; otherwise, it becomes "unethical to ask patients to provide feedback which will never be taken into account" [31]. Thus, service providers should ensure that routinely collected feedback (including after an Attend Anywhere appointment) is meaningful for both patients and clinicians, serves a beneficial function beyond mandatory feedback collection, and focuses on care delivery aspects that are most important to patients and clinicians. Future research may consider which feedback

methods are most effective in encouraging responses, particularly in the new *digital norm*, with staff members supported and empowered in acting upon and responding to feedback received.

Limitations

The first limitation of this research is the reliance on a self-selected sample of individuals who attended or facilitated a VC and chose to provide feedback. Experiences or barriers for those unable or choosing not to use VCs currently remain unknown, as do the experiences of those not providing feedback following their VC. Nevertheless, from this study, we know that hundreds of staff members and patients had a positive experience with VCs. Second, owing to the anonymous nature of the data set, we were unable to identify how many individuals completed the survey on repeated occasions; thus, the results may be skewed by repeat respondents. Third, this research relies on secondary data and the subsequent questions or scales used by the Trust. Although informed by the *1-week implementation guide for NHS Trusts and NHS Foundation Trusts* provided by NHS England and NHS Improvement [23], the survey had some limitations—some of the questions were poorly worded, and the questions were not directly comparable between staff and patient questionnaires (a more detailed discussion is provided in Multimedia Appendix 1). Furthermore, to the researchers' knowledge, the survey was not created in co-design with all relevant stakeholders, including patients and carers. Therefore, the questions asked may not reflect the most important aspects of the VC experience for patients.

Future research could explore challenges and barriers for excluded patients, particularly those considered seldom heard or marginalized in the context of digital health [32]. Murphy et al [12] noted previously that digital consultations increase access for those with information technology skills but may reinforce existing health inequalities. There is an important balance required between acknowledging increased accessibility for some patient groups who may encounter difficulties accessing face-to-face services and acknowledging reduced accessibility for others. This aspect needs urgent further work and reiterates the importance of patient choice and availability of multiple media to access health and care services.

The nature of this study meant that we were also unable to conclude on a number of additional factors, highlighting the scope for further research. Apparent efficiency savings in service delivery via VC should be explored to assess the impact of VCs on clinician and patient experience. Although intensive work may aid in meeting growing health care requirements, workforce burnout poses a danger. Related to efficiency, further research should look to establish DNAs before and since the implementation of VCs, with almost a quarter of staff reporting less DNAs using VCs than with usual practice. This is an interesting result, and further work could explore the reasons for the reductions in DNAs compared with in-person consultations. Although the results suggest a reduction in DNAs generally, 3 staff comments suggested increases. It would be interesting to explore if DNAs that do occur are linked to patient inability to access VCs or lack of confidence with technology. Other implications for future research include a need to identify

the mechanisms responsible for the positive patient experiences and high levels of future VC use intentions, as demonstrated by the Trust in this research. By doing so, other Trusts and health care services can engage with acknowledged areas of best practice. Economic evaluations that incorporate clinician, patient, and environmental savings may also be beneficial, although it is important to emphasize that potential cost savings should not take precedence over patient safety, quality of care, and stakeholder experience.

In addition, suggestions made in this research that VCs improve "affordability" of appointments and comfort in sharing personal or clinical information are important areas of future interest. Future research questions could include how and in what ways have VCs affected patient accessibility? Similarly, how, if at all, do VCs affect the therapeutic relationship? Future research could compare patient satisfaction with more conventional face-to-face consultations or other VC platforms. In addition, the range and type of consultations available to patients are currently limited and expressed satisfaction may reflect a lack of choice or alternatives. Future research may review patient experiences over time, particularly during times of heightened and reduced COVID-19 restrictions.

Comparison With Previous Work

Generally, our results continue to demonstrate positive experiences for staff and patients with VCs during the pandemic [4,9,10,16], furthering previous work with smaller samples and narrower focus. For example, congruent with Gilbert et al [4], we found that positively perceived aspects of VCs included reduced travel times and reduced impact of travel on symptoms such as fatigue [4]. Other similarities with previous work include low confidence reported among some participants [4] and the negative impact of technological limitations and difficulties on patient experience [4,9,12,13]. Findings from this research regarding age differences in independent use and family involvement are also congruent with other research [12,33]. However, given the difficulties that many older adults have in traveling to outpatient clinics [34] and the largely high acceptability of VC use reported in this study for older adults, no quick assumptions should be made about the unsuitability of VC for older adults.

Areas of divergence from the existing literature include patients reporting higher levels of satisfaction and willingness to use VCs in the future than in previous work [4]. Previous feedback was collected within an entirely orthopedic service, which could suggest that greater satisfaction and use intentions are seen here owing to the variety of services included, which may better

translate to VC than orthopedics as perhaps a more *hands-on* service. However, this would need further exploration as survey limitations impair our understanding of exactly which service patient respondents accessed. Other contributions of this research include the identification of additional benefits, including enhanced comfort and subsequent ability to "open up more." This contrasts with the results of Liberati et al [14], who reported impairments to depth of conversation and relational quality via remote means. Although this paper reports on a larger sample, Liberati et al [14] also reported on qualitative interviews. Therefore, this incongruence in results is worth exploring further, perhaps across specific psychological therapy services. In contrast to Isautier et al [9], who suggested that telehealth limitations included poorer quality of communication, our results suggest that most patients were satisfied with the VC aspects related to communication, with combined technical satisfaction being lower, congruent with Kayser et al [10]. In addition, in this study, we provide further insight into the influence of patient age and device used in predicting overall VC experience, with implications for targeted consultations in the future. This research also provides interesting insights into both staff and patients reporting an increased sense of accessibility and patient perception of enhanced affordability. Therefore, our work contributes to furthering previous research [4,9-17] that reported on small sample sizes and generally single-service focus, whereas we report on a comparatively large sample across outpatient services in a rural and older adult setting [18,20].

Conclusions

In conclusion, most NHS staff members and patients reported positive experiences with VCs for outpatient care in a rural, older adult, and deprived setting. Patients often felt listened to, able to communicate their needs, and understood, and staff and patients noted resource savings and enhanced accessibility. However, some barriers identified, such as technological difficulties, accessibility of patient information, and accessibility or suitability concerns, require further attention if the potential benefits of VCs are to be realized and their use is to be sustained. The implications of this research include the implementation of patient-suggested improvements, including trial calls, turning music off, facilitating photo uploads, expanding written character limit, and supporting VCs on other browsers. Future work may explore the accessibility and experience of patients excluded from this study through lack of VC access. In addition, this study demonstrated the real-world impact of routine feedback and raises further discussion on the future use of routine staff and patient experience data.

Authors' Contributions

All authors have read and approved the manuscript. HB analyzed and interpreted the results and led in producing the manuscript. RB analyzed and interpreted the results and substantively contributed to the manuscript. KE analyzed and interpreted the results

and substantively contributed to the manuscript. SS reviewed and contributed to the manuscript. KA provided expertise in patient experience and engagement and reviewed and substantively contributed to the manuscript. EW facilitated access to patient experience data, provided contextual data, and reviewed and substantively contributed to the manuscript. AC reviewed and substantively contributed to the manuscript. RJ analyzed and interpreted the results and reviewed and substantively contributed to the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questions and response options.
[DOCX File , 24 KB - formative_v6i3e30486_app1.docx]

References

1. Aprahamian I, Cesari M. Geriatric syndromes and SARS-Cov-2: more than just being old. J Frailty Aging 2020;9(3):127-129 [FREE Full text] [doi: 10.14283/jfa.2020.17] [Medline: 32588023]

2. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]

3. Ohannessian R, Duong TA, Odone A. Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: a call to action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]

4. Gilbert AW, Billany JC, Adam R, Martin L, Tobin R, Bagdai S, et al. Rapid implementation of virtual clinics due to COVID-19: report and early evaluation of a quality improvement initiative. BMJ Open Qual 2020 May 21;9(2):e000985 [FREE Full text] [doi: 10.1136/bmjoq-2020-000985] [Medline: 32439740]

5. Hollander JE, Carr BG. Virtually perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]

6. Iyengar K, Jain VK, Vaishya R. Pitfalls in telemedicine consultations in the era of COVID 19 and how to avoid them. Diabetes Metab Syndr 2020;14(5):797-799 [FREE Full text] [doi: 10.1016/j.dsx.2020.06.007] [Medline: 32534432]

7. Greenhalgh T, Shaw S, Seuren L, Wherton J. Video consultation information for NHS Trusts and Foundation Trusts. University of Oxford. URL: https://www.england.nhs.uk/coronavirus/wp-content/uploads/sites/52/2020/08/C0638-nhs-vc-info-for-nhs-trusts.pdf [accessed 2022-02-21]

8. Remote consultations guidance under COVID-19 restrictions. Royal College of Nursing. URL: https://www.rcn.org.uk/professional-development/publications/rcn-remote-consultations-guidance-under-covid-19-restrictions-pub-009256 [accessed 2022-02-22]

9. Isautier JM, Copp T, Ayre J, Cvejic E, Meyerowitz-Katz G, Batcup C, et al. People's experiences and satisfaction with telehealth during the COVID-19 pandemic in Australia: cross-sectional survey study. J Med Internet Res 2020 Dec 10;22(12):e24531 [FREE Full text] [doi: 10.2196/24531] [Medline: 33156806]

10. Kayser M, Valtin C, Greer M, Karow B, Fuge J, Gottlieb J. Video consultation during the COVID-19 pandemic: a single center's experience with lung transplant recipients. Telemedicine e-Health 2020 (forthcoming). [doi: 10.1101/2020.05.12.20093799]

11. Barsom EZ, Feenstra TM, Bemelman WA, Bonjer JH, Schijven MP. Coping with COVID-19: scaling up virtual care to standard practice. Nat Med 2020 May;26(5):632-634. [doi: 10.1038/s41591-020-0845-0] [Medline: 32405054]

12. Murphy M, Scott LJ, Salisbury C, Turner A, Scott A, Denholm R, et al. Implementation of remote consulting in UK primary care following the COVID-19 pandemic: a mixed-methods longitudinal study. Br J Gen Pract 2021 Jan 17;71(704):e166-e177. [doi: 10.3399/bjgp.2020.0948]

13. Bhardwaj A, Moore A, Cardinal RN, Bradley C, Cross L, Ford TJ. Survey of CAMHS clinicians about their experience of remote consultation: brief report. BJPsych Open 2021 Jan 13;7(1):e34 [FREE Full text] [doi: 10.1192/bjo.2020.160] [Medline: 33436136]

14. Liberati E, Richards N, Parker J, Willars J, Scott D, Boydell N, et al. Remote care for mental health: qualitative study with service users, carers and staff during the COVID-19 pandemic. BMJ Open 2021 Apr 22;11(4):e049210 [FREE Full text] [doi: 10.1136/bmjopen-2021-049210] [Medline: 33888531]

15. Dhahri A, Iqbal M, Pardoe H. Agile application of video telemedicine during the COVID-19 pandemic. Cureus 2020 Nov 04;12(11):e11320 [FREE Full text] [doi: 10.7759/cureus.11320] [Medline: 33262918]

16. Doraiswamy S, Abraham A, Mamtani R, Cheema S. Use of telehealth during the COVID-19 pandemic: scoping review. J Med Internet Res 2020 Dec 01;22(12):e24087 [FREE Full text] [doi: 10.2196/24087] [Medline: 33147166]

17. Vusirikala A, Ensor D, Asokan AK, Lee AJ, Ray R, Tsekes D, et al. Hello, can you hear me? Orthopaedic clinic telephone consultations in the COVID-19 era- a patient and clinician perspective. World J Orthop 2021 Jan 18;12(1):24-34 [FREE Full text] [doi: 10.5312/wjo.v12.i1.24] [Medline: 33520679]

XSL•FO
RenderX

18. How old is the UK? An investigation into the ageing population of Britain. Age UK. URL: https://www.ageukmobility.co.uk/mobility-news/article/how-old-is-the-uk [accessed 2022-02-22]

19. Cornwall: a brief description. Cornwall Council. URL: https://www.cornwall.gov.uk/media/20392018/cornwall-statistics-infographic-a3_proof3.pdf [accessed 2022-02-22]

20. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. BMJ Qual Saf 2016 Dec 14;25(12):986-992 [FREE Full text] [doi: 10.1136/bmjqs-2015-004411] [Medline: 26369893]

21. Attend anywhere. NHS Norfolk Community Health and Care. URL: https://www.norfolkcommunityhealthandcare.nhs.uk/attend-anywhere/ [accessed 2022-02-22]

22. Specialty guides for patient management. NHS England and NHS improvement coronavirus. URL: https://www.nric.org.uk/node/54267 [accessed 2022-02-22]

23. 'Attend Anywhere' Patient Feedback _ ESNEFT. NHS Foundation Trust. URL: https://forms.office.com/Pages/ResponsePage.aspx?id=kp4VA8ZyI0umSq9Q55Ctvy0O6tkePahBgNX4hXWu7AVUQVJNQ1g1REM4UFQwNDczUDUwTU9NTUc1My4u [accessed 2022-02-22]

24. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

25. McGaughey RE, Zeltmann SM, McMurtrey ME. Motivations and obstacles to smartphone use by the elderly: developing a research framework. Int J Electron Finance 2013;7(3/4):177. [doi: 10.1504/ijef.2013.058601]

26. Vaportzis E, Clausen MG, Gow AJ. Older adults perceptions of technology and barriers to interacting with tablet computers: a focus group study. Front Psychol 2017 Oct 04;8:1687 [FREE Full text] [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]

27. Biswas M, Romeo M, Cangelosi A, Jones RB. Are older people any different from younger people in the way they want to interact with robots? Scenario based survey. J Multimodal User Interfaces 2019 Jul 24;14(1):61-72. [doi: 10.1007/s12193-019-00306-x]

28. Sibley M, Earwicker R, Huber JW. Making best use of patient experience. J Clin Nurs 2018 Dec;27(23-24):4239-4241. [doi: 10.1111/jocn.14504] [Medline: 29752838]

29. Edwards C, Staniszewska S. Accessing the user's perspective. Health Soc Care Community 2000 Nov;8(6):417-424. [doi: 10.1046/j.1365-2524.2000.00267.x] [Medline: 11560712]

30. Williams B, Coyle J, Healy D. The meaning of patient satisfaction: an explanation of high reported levels. Social Sci Med 1998 Nov;47(9):1351-1359. [doi: 10.1016/s0277-9536(98)00213-5]

31. Sheard L, Peacock R, Marsh C, Lawton R. What's the problem with patient experience feedback? A macro and micro understanding, based on findings from a three-site UK qualitative study. Health Expect 2019 Feb 22;22(1):46-53 [FREE Full text] [doi: 10.1111/hex.12829] [Medline: 30244499]

32. Burt J, Campbell J, Abel G, Aboulghate A, Ahmed F, Asprey A, et al. Improving patient experience in primary care: a multimethod programme of research on the measurement and improvement of patient experience. Programme Grants Appl Res 2017 Apr;5(9):1-452. [doi: 10.3310/pgfar05090] [Medline: 28654227]

33. Antoni E, Jaboli M, Samarasinghe M, Mohamed A, Epstein O. PWE-132 Is the NHS ready for video-conference consultations in outpatient clinics? Gut 2016 Aug 17;65(Suppl 1):A203-A204. [doi: 10.1136/gutjnl-2016-312388.377]

34. Painful journeys: why getting to hospital appointments is a major issue for older people. Age UK. URL: https://www.ageuk.org.uk/globalassets/age-uk/documents/reports-and-publications/reports-and-briefings/active-communities/rb_dec17_painful_journeys_indepth_report.pdf [accessed 2022-02-22]

Abbreviations

DNA: did not attend
NHS: National Health Service
VC: video consultation

Original Paper

The Effect of Noise-Masking Earbuds (SleepBuds) on Reported Sleep Quality and Tension in Health Care Shift Workers: Prospective Single-Subject Design Study

Nicole M Duggan[1], MD; M Adrian Hasdianda[1], MD, MSc; Olesya Baker[2], PhD; Guruprasad Jambaulikar[1], MPH, MBBS; Andrew J Goldsmith[1], MD, MBA; Anna Condella[1], MD; Desiree Azizoddin[3], PsyD; Adaira I Landry[1], MD, MEd; Edward W Boyer[1], MD, PhD; Andrew J Eyre[1], MD, MEd

[1]Department of Emergency Medicine, Brigham and Women's Hospital, Boston, MA, United States

[2]Center for Clinical Investigation, Brigham and Women's Hospital, Boston, MA, United States

[3]Health Promotion Research Center, Department of Family and Preventive Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK, United States

Corresponding Author:
Nicole M Duggan, MD
Department of Emergency Medicine
Brigham and Women's Hospital
75 Francis Street
NH-2
Boston, MA, 02115
United States
Phone: 1 617 724 4068
Email: nmduggan@partners.org

Abstract

Background: Shift work is associated with sleep disorders, which impair alertness and increase risk of chronic physical and mental health disease. In health care workers, shift work and its associated sleep loss decrease provider wellness and can compromise patient care. Pharmacological sleep aids or substances such as alcohol are often used to improve sleep with variable effects on health and well-being.

Objective: We tested whether use of noise-masking earbuds can improve reported sleep quality, sleepiness, and stress level in health care shift workers, and increase alertness and reaction time post night shift.

Methods: Emergency medicine resident physicians were recruited for a prospective, single-subject design study. Entrance surveys on current sleep habits were completed. For 14 days, participants completed daily surveys reporting sleep aid use and self-rated perceived sleepiness, tension level, and last nights' sleep quality using an 8-point Likert scale. After overnight shifts, 3-minute psychomotor vigilance tests (PVT) measuring reaction time were completed. At the end of 14 days, participants were provided noise-masking earbuds, which they used in addition to their baseline sleep regimens as they were needed for sleep for the remainder of the study period. Daily sleep surveys, post–overnight shift PVT, and earbud use data were collected for an additional 14 days. A linear mixed effects regression model was used to assess changes in the pre- and postintervention outcomes with participants serving as their own controls.

Results: In total, 36 residents were recruited, of whom 26 participants who completed daily sleep surveys and used earbuds at least once during the study period were included in the final analysis. The median number of days of earbud use was 5 (IQR 2-9) days of the available 14 days. On days when residents reported earbud use, previous nights' sleep quality increased by 0.5 points ($P<.001$, 95% CI 0.23-0.80), daily sleepiness decreased by 0.6 points ($P<.001$, 95% CI –0.90 to –0.34), and total daily tension decreased by 0.6 points ($P<.001$, 95% CI –0.81 to –0.32). These effects were more pronounced in participants who reported worse-than-average preintervention sleep scores.

Conclusions: Nonpharmacological noise-masking interventions such as earbuds may improve daily sleepiness, tension, and perceived sleep quality in health care shift workers. Larger-scale studies are needed to determine this interventions' effect on other populations of shift workers' post–night shift alertness, users' long-term physical and mental health, and patient outcomes.

XSL·FO
RenderX

Introduction

Over the last several decades, sleep deficiency has reached epidemic proportions. An estimated one-third of adults in the United States do not get the recommended seven hours of sleep nightly, and at least 50 million Americans have chronic sleep disorders [1,2]. Shift work, including working overnight hours, having variable or rotating schedules, or working extended hours on call uniquely contributes to worse sleep and poor work-related outcomes [3-6]. Health care practitioners and support staff represent occupation groups with some of the highest prevalence of shift work and sleep problems including sleep disturbances (eg, multiple awakenings), and sleep deficiency from either extended wake episodes (ie, acute sleep deprivation) or multiple days or nights of insufficient sleep (ie, sleep restriction or chronic sleep deprivation) [7].

In health care workers, sleep loss impairs alertness, which can lead to poor work performance and medical errors, potentially compromising patient care [4,8-11]. Despite the recognized impact of sleep loss on clinician health and patient safety, workplace interventions to combat sleep loss and improve restful sleep are lacking compared to other wellness initiatives [12]. Accordingly, we assessed the impact of noise-masking earbuds on self-reported sleepiness, sleep quality, stress level, and post–overnight shift alertness in emergency medicine (EM) resident physicians.

Methods

Study Design and Setting

This prospective, single-subject design study was performed at a single urban academic medical center. This facility hosts residency training programs for multiple specialties including a 4-year EM training program.

Ethical Considerations

This work was approved by the institutional review board (IRB), and all participants gave informed consent. This study was approved by the Brigham and Women's Hospital Institutional Review Board (2019p002509).

Participant Recruitment

EM resident physicians working full-time at the emergency department (ED) during the study period were recruited. All EM residents who were not hearing impaired and who owned a smart phone capable of receiving SMS text messages and opening electronic surveys were eligible for inclusion. Participants were recruited via email to the residency listserv and through a text message in a resident-specific group messaging app delivered to their personal devices.

Within this residency training program, each resident works between 18-21 shifts in a rotating schedule comprising day shifts (roughly 7 AM to 5 PM), twilight shifts (roughly 3 PM to 2 AM), or overnight shifts (roughly 11 PM to 8 AM) over each 28-day period. Each participant works 5-7 overnight shifts over a 28-day period.

Interventions

Following consent procedures, participants completed an electronic entrance survey that included questions about their baseline sleep habits, sleep aid use, self-reported sleep quality, daily sleepiness, and daily tension over the prior 28 days (Multimedia Appendix 1). Over the next 14 days (the control period), participants were instructed to continue with their baseline sleep habits. On study day 15, each participant was provided with a pair of Bose SleepBuds (Bose Corp). SleepBuds (hereby referred to as "earbuds") are earbuds worn in both ears all night and function to mask ambient noises by playing various sound tracks selectable from the Bose app, which each target decibels that are specific to common ambient sounds. Participants were given instruction on device functionality including the app for activating the device noise-masking technology, which only occurs when the device is turned on and used in conjunction with the app. Participants were instructed to use the earbuds with the noise-masking technology for their sleep episodes as needed over the next 14 days in addition to their preferred baseline sleep aids. This 14-day period was considered the intervention period. Earbud use was at the discretion of each participant, and study personnel who were not affiliated with the Bose Corporation were available throughout the study period to answer questions about functionality or troubleshoot device function. At the end of the study period, participants completed an electronic exit survey on sleep aid use and earbud functionality over the prior 28-day period (Multimedia Appendix 1).

Measurements

Daily Surveys

Each day of the 28-day study period, participants received an automated SMS text message at 12 noon containing a daily survey asking them to rate the quality of their last sleep episode, current daytime sleepiness, and current level of tension. All measures were reported on 8-point Likert scales of 0 (extremely bad sleep quality, not sleepy at all, and not tense at all) to 7 (extremely good sleep quality, extremely sleepy, and extremely tense). Beginning on day 15, daily surveys also included a question about whether earbuds were used during the last sleep episode (Multimedia Appendix 2).

Psychomotor Vigilance Testing

After every overnight shift, each participant was approached by a trained study research assistant to complete a 3-minute psychomotor vigilance test (PVT, PVT Research Tool, Texas A&M University System CSE) [13]. This test was conducted on portable tablet devices requiring participants to rapidly tap a circle as it appeared on a tablet screen, and the reaction time for each tap was measured. Participants were able to decline

participation in the daily PVT if clinical demands required their attention.

6-Month Follow-up

After the study period, participants were allowed to keep their earbuds for their own personal use. At 6 months after study completion, participants were sent an electronic survey asking about sleep aid use since study completion, including earbud use. For participants still using earbuds, additional questions about when and how they most commonly used the device since the study period were included.

Data Management and Statistical Analyses

Study data were collected and managed using REDCap electronic data capture system sponsored and hosted at our institution. Data were deidentified prior to analysis and analyzed using Stata (version 16; StataCorp). Frequencies and percentages were used to summarize binary and categorical variables. Mean (SD) or median (IQR) values were used to summarize time variables and Likert scale data.

Changes in quality of the last sleep episode, current daytime sleepiness, and current level of tension in control and intervention periods were assessed using a linear mixed effects regression model with a participant-specific random intercept. In the postimplementation period, we only included data collected on days when participants reported using the earbuds. For PVT data, each participant's mean reaction times pre- and postintervention period were directly compared. Six-month follow up responses were qualitatively assessed.

Results

Participant Demographic Characteristics

Of 58 invited participants, 38 participants enrolled in the study. Of these, 6 participants who did not complete any daily surveys and 6 who participated in surveys but never used the earbuds were excluded. The remaining 26 participants completed a total of 655 daily sleep surveys. We further excluded surveys taken on postintervention days when participants reported not using earbuds, resulting in 501 daily surveys being included in final analysis (Figure 1). Participants included in final analysis represented resident physicians from all 4 postgraduate years. In total, 27% of participants included in the final analysis were female, and their ages ranged from 25 to 35 years (Table 1).

Figure 1. Study design flow chart. PVT: psychomotor vigilance test.
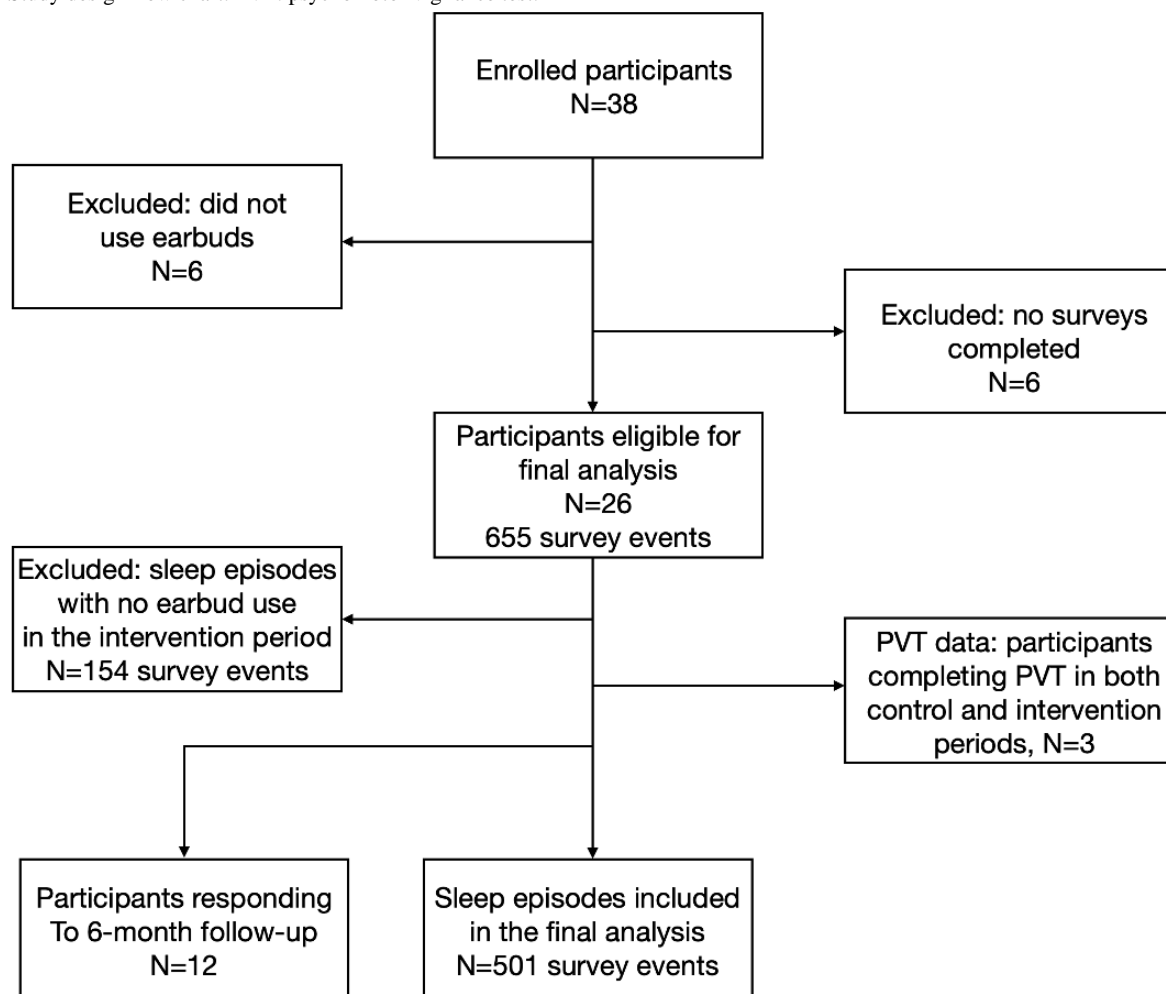
Table 1. Participant demographic characteristics (N=26).

Variables	Participants
Sex, n (%)	
Male	19 (73.1)
Female	7 (26.9)
Age (years), n (%)	
20-25	0 (0)
25-30	12 (46.1)
30-35	14 (53.9)
>35	0 (0)
Training year, n (%)	
PGY1	5 (19.2)
PGY2	9 (34.6)
PGY3	7 (26.9)
PGY4	5 (19.2)
Baseline average sleep or tension measures over the 4 prior weeks, median (IQR)	
Sleep quality	5 (4-6)
Sleepiness	4.5 (4-5)
Tension	5 (3-5)
Baseline sleep and aid use, n (%)	
Blackout curtains	16 (61.5)
Eye mask	8 (30.8)
Earplugs	6 (23.1)
Weighted blanket	1 (3.8)
White noise	9 (34.6)
Pharmacological sleep aid[a]	9 (34.6)
Other	1 (3.8)
None of the above	2 (7.7)

[a]Pharmacological sleep aid refers to oral medications such as melatonin, benzodiazepines, zolpidem, or diphenhydramine.

Earbud Use

Among participants who reported using earbuds at least once, the frequency of use varied. Of 14 days when earbuds were available, participants used earbuds for a median of 5 (IQR 2-9) days. The lowest earbud use occurred on day one of the intervention, with 6 of 26 participants reporting use, trending upward until day 4 when 15 of 26 participants reported using earbuds, and remaining generally stable with a slight downward trend in use toward the end of the study (Figure 2). Participants who reported using earbuds for ≥7 of the 14 days did not have significantly different baseline prestudy 4-week sleep scores than those who reported using earbuds between 1 and 6 days (mean score 5.2 vs 4.8, *P*=.34).

**Figure 2.** Reported earbud use over the intervention period.
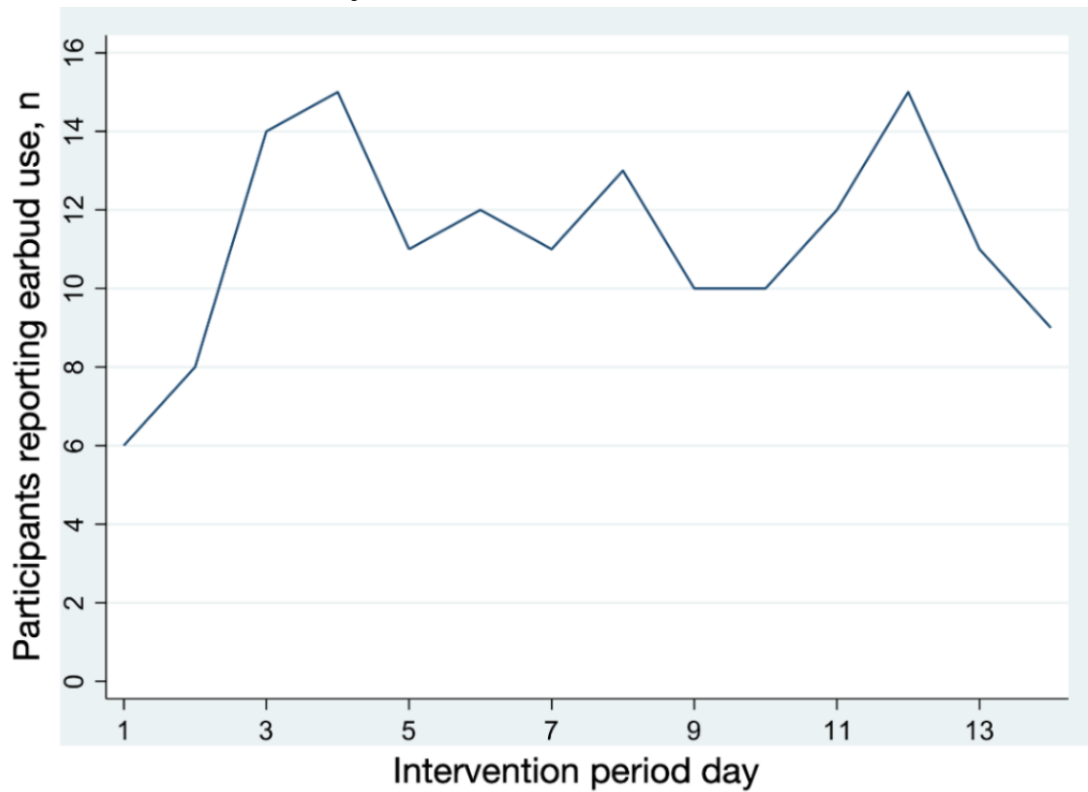
Sleep Quality, Daily Sleepiness, and Tension

Participants' self-reported last sleep episode quality, daily sleepiness, and daily tension had improving trends in the postintervention period (Figures 3A-C). On days when residents used earbuds, on Likert scales of 0 to 7 points, previous nights' sleep quality increased by 0.5 points ($P<.001$, 95% CI 0.23-0.80), daily sleepiness decreased by 0.6 points ($P<.001$, 95% CI –0.90 to –0.34), and total daily tension decreased by 0.5 points ($P<.001$, 95% CI –0.81 to –0.32) using linear mixed regression models. In the subset of participants who reported below-median sleep scores before the intervention, the beneficial effects of using earbuds were amplified.

Figure 3. Self-rated last nights' sleep quality (A), daily sleepiness (B), and daily tension (C) on days with earbud use during the prior sleep period.
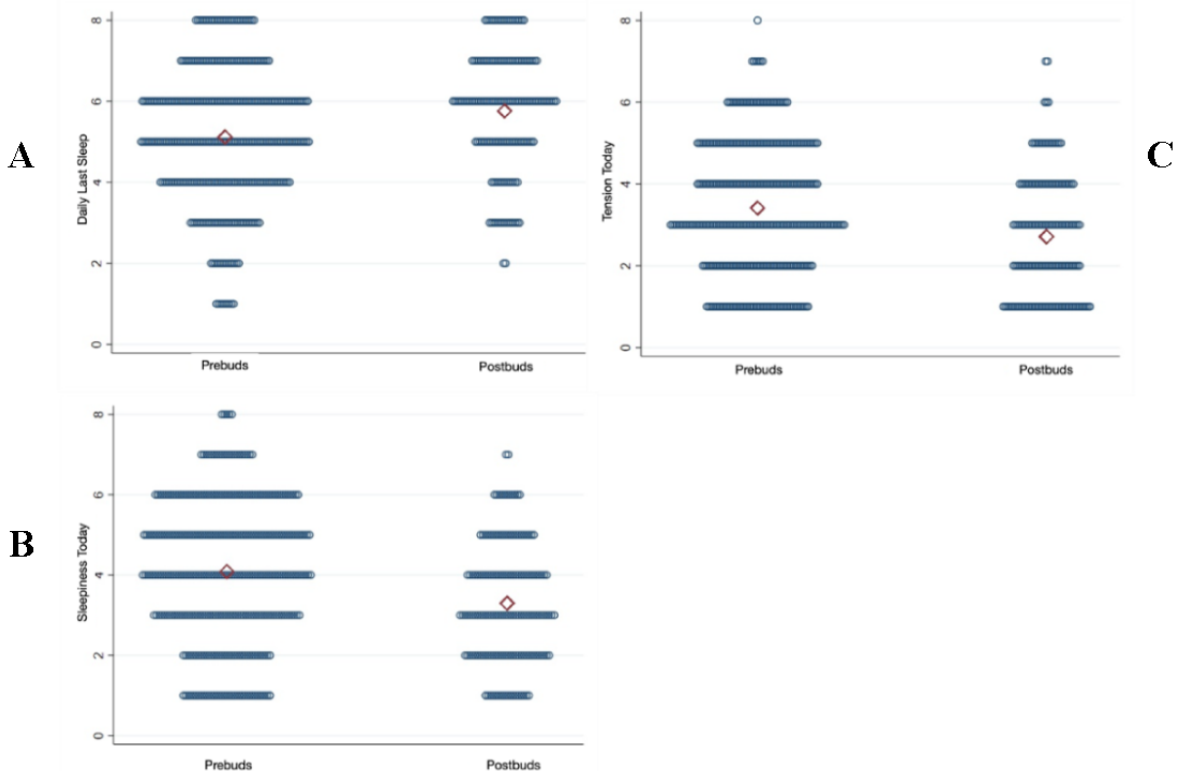
Post–Night Shift Reaction Time and Alertness

In total, 7 post–overnight shift residents completed a total of 12 PVT tests during the intervention period. A total of 6 completed PVTs were linked to participants who reported no earbud use and were thus excluded from the final analysis. Ultimately, 3 participants both used the earbuds and completed PVTs in both the control and intervention periods. Data from these participants comparing pre- and postearbuds mean reaction time showed no significant differences; however, this study was underpowered to detect an effect of earbud use on PVT.

6-Month Follow-up

We received responses from 12 (46%) participants. Of the 12 respondents, 5 reported ongoing earbud use; all of whom reported use on less than 25% of all sleep episodes since study completion. Participants cited limited sound options, uncomfortable fit, and forgetting to use the device as barriers to further use. There were no notable differences in self-reported baseline sleep aid use and 6-month follow up sleep aid use.

Discussion

Principal Findings

We report that adding a noise-masking earbud, Bose SleepBuds, to a sleep routine can foster significant improvements in shift workers' self-reported sleep quality, sleepiness, and tension over baseline sleep habits. Participants in this sample noted statistically significant improvements following use of earbuds, and minimal safety concerns were identified. In health care providers, poor sleep contributes to decreased alertness and performance, which adversely affects patient care. Sleep-deprived shift workers demonstrate decreased mood, and professional fulfillment, emotional exhaustion, and increased rated of burnout [10,14-18].

Sleep and wake disturbances are common among shift workers and can produce a primary circadian rhythm disorder called "shiftwork sleep disorder" (SWSD) characterized by insomnia, excessive sleepiness, and significant sleep loss [6,19,20]. Sleep problems are associated with debilitating chronic physical and mental health disease. In addition to excessive sleepiness, sleep deficiency and SWSD are associated with critical short- and long-term health concerns. Sleep deficiency in shift workers has been associated with obesity, type 2 diabetes, stroke, multiple types of cancers, and cardiovascular disease [21-24].

To combat shift work–associated sleep loss, shift workers may develop self-designed strategies to assist with sleep. One common approach is to alternate between the use of pharmacological stimulants and sedatives to achieve wakefulness or sleep as needed. Staying awake and alert during shift can require aids such as caffeine or nicotine [25,26]. To accelerate the onset of sleep, over-the-counter supplements such as melatonin, prescription sleep aids such as benzodiazepines, or other sedative-hypnotic medications (eg, so-called "Z-drugs" such as zolpidem) are also often used [26-28]. Despite the fact that alcohol consumption worsens sleep overall, postshift alcohol use is another common strategy among shift workers to achieve sleep, a strategy that has been associated with binge drinking disorder [29,30]. Each of these strategies demonstrate variable efficacy and can have significant off-target, often deleterious, health effects, development of compulsive use consistent with addiction, and poor patient care.

Participant adherence and uptake in the intervention was one challenge faced in this study. When considering wellness interventions, in addition to intervention efficacy, institutions must also consider factors such as participant interest and adherence and cost-benefit ratios of a given intervention. Here, roughly two-thirds of the approached participants elected to enroll in the study in which they were receiving sleep aid devices free of charge. It is possible this number may serve as a proxy for interest in such nonpharmacological sleep aid devices or in sleep interventions in general. Direct assessment of shift worker interest in nonpharmacological sleep interventions, such as this one, are needed. Interestingly, a larger proportion of males than females enrolled in the study despite the fact that the group originally solicited was roughly even in terms of sex. This may suggest that an electronic sleep device such as noise-masking earbuds are more attractive to males than to females as an intervention. Further analysis of the most likely users may be helpful in targeting sleep interventions to users who may benefit from or adhere most to them. Finally, cost-benefit ratio is often a consideration when implementing wellness initiatives. Though the particular noise-masking earbuds studied here have a relatively high price point for the general public, targeting these devices to participants who may have the highest uptake, adherence, and physiological outcomes may result in a favorable cost-benefit ratio for organizations. Further work is needed to more carefully target sleep interventions such as the earbuds studied here.

Shift workers experience striking rates of workplace burnout; yet, occupational interventions targeting sleep problems are often lacking compared to other wellness initiatives [12]. The importance of this work, therefore, is that noise-masking earbuds may improve several metrics of sleep quality without the threat of problematic substance use, medication adverse events, or pharmaceutical misuse. Although we identified several beneficial effects of the noise-masking earbuds on self-reported sleep and tension, alertness testing via PVT was underpowered to detect an effect in our study population. Because health care professionals are particularly vulnerable to the deleterious effects of sleep loss and SWSD is prevalent in up to 30% of the general population, our findings, while limited, have broad therapeutic potential in multiple industries where shift work is common, including medicine.

Limitations

This was a single-arm study with a small sample size, and variable intervention adherence, all of which affected the analytical power of this study. Further efficacy should be evaluated in a randomized controlled trial comparing earbud use to a matched control. Here, we focused on resident physicians as representatives of the health care shift worker population. This population tends to be fairly young (between 25-35 years old), and may be more willing to use new technology than shift workers as a whole. While approximately one-third of resident shifts are overnight, night shifts tend to be clustered and are not evenly distributed over a 28-day period.

Thus, typical resident schedules are somewhat suboptimal for a single-subject design where subjects serve as their own controls across 28-days. The PVT data collection was limited given the smaller number of data points across control and intervention periods for individual participants. Future solutions may include recruiting study participants such as nocturnists, who are physicians who work a higher proportion and more consistent distribution of night shifts, other groups who work exclusively overnight, or participants with clinically diagnosed SWSD. Increasing study population size or extending the study period by multiple months would likely enhance our data points and improve study power.

Variable intervention adherence also affected our findings. Manually filling out daily surveys and PVT testing after overnight shifts can be burdensome to participants. Participants may have forgotten to complete surveys after receiving the SMS text message reminder. Future studies will consider alternative data collection strategies that may present less of a barrier, such as automatic data collection on sleep quality via wearable fitness trackers or cellular telephones. These strategies may also provide more objective measures of sleep quality as opposed to self-reporting. In terms of inconsistent device use, participants cited a desire for additional sound options including selectable music, improved device fit, and needing reliable reminders to improve intervention adherence. Addressing these factors will be crucial to designing maximally effective digital therapies for sleep deprivation in the future.

The potential for sustained adherence is an important consideration in any wellness intervention. Here, at 6-month follow up, <50% of respondents reported ongoing earbud use beyond the study period. Six-month follow-up responses were collected from only 46% of original participants; thus, the limited response rate likely impacts our ability to draw firm conclusions from these data. Subjectively, however, reasons commonly cited for limited ongoing use included poor fit, limited audio tracks, and forgetting to use the earbuds. The fit and audio selection can be addressed by manufacturers to make such products more attractive to users. Habit formation and establishing automated behaviors are key to implementing sustainable health and wellness interventions. Data from nutritional wellness interventions suggest habits can take more than 50 days to reach automaticity [31]. Regarding forgetting to use the device, it is likely our study period was not long enough to promote habit formation. Prolonging the intervention study period with frequent reminders to use the device, or instructing participants to place the device in a convenient, easily accessible location that will remind and reinforce use can be helpful. Further, time-based cues such as reminders on cellular telephones can help establish behaviors while habit formation is taking place [31].

As a pilot study, the aim of this work was to assess uptake, user receptiveness, and perceived effect of a nonpharmacological sleep intervention. Important next steps for future studies include evaluating objective measures of sleep and mood, physician burnout among earbud users, as well as potential changes in clinician performance and patient outcomes such as physician errors or return visit rates. While participant experience is an important component in evaluating the effect of a wellness intervention, objective outcomes studies will help corroborate the utility of implementing such interventions on an institutional level.

Comparison With Prior Work

While electronic devices aimed at addressing sleep problems do currently exist, few have robust data supporting their efficacy. Most common devices addressing poor sleep include fitness trackers or mobile apps to track sleep quality, but these rarely offer therapeutic interventions [32,33]. Existing digital therapies for poor sleep include app-based meditation or relaxation guides, and app-guided cognitive behavioral therapy [34,35]. Though moderately effective in some populations, these interventions are often targeted at addressing chronic sleep loss and insomnia, are subject to intervention fatigue, and do not address acute sleep problems and circadian rhythm disruption common in post–night shift health care workers.

Devices supplying white noise are commonly marketed as sleep aids. Despite broad commercial penetrance, evidence for the efficacy of continuous white noise on sleep onset, latency, or quality is inconsistent [36-39]. Many factors likely play into the effect of ambient sounds on sleep, including but certainly not limited to the type of sound, the ability to block out alternative ambient noises, and sound volume among other factors. The SleepBuds technology used in this study included first-generation devices that provide noise-masking technology to reduce ambient auditory distractions with optional relaxing sounds. Given our promising results, future work can explore the effect of subsequent-generation devices with enhanced functionality targeted at relaxation. This work will both assess the efficacy of this digital intervention, as well as contribute to the existing body of work on the therapeutic effect of ambient noise on sleep overall.

Conclusions

Sleep loss is a public health crisis that disproportionately affects the physical and mental health of shift workers. The recent emphasis on support programs aimed at improving employee wellness and decreasing burnout is not reflected in interventions aimed at improving sleep. Our data demonstrate the efficacy of a digital intervention addressing sleep disturbance. This intervention is of low risk and is feasible in a complex cohort such as medical resident physicians. While our target population was health care shift workers, extension of our findings to other industries and cohorts is reasonable. Our work prepares for additional studies to improve employee health and wellness through effective, generalizable, and scalable sleep interventions.

Acknowledgments

Authors' Contributions

NMD conceptualized and designed the study, wrote and submitted the IRB draft, acquired, analyzed, and interpreted the data, and drafted and revised the manuscript. MAH and GJ oversaw IRB draft submission and data acquisition and management. OB carried out data analysis and interpretation and manuscript drafting and revision. AJG drafted and revised the manuscript. AC and AIL conceptualized and designed the study, wrote and prepared the IRB draft, and revised the manuscript. DA mentored the design of the study and revised the manuscript. EWB mentored the design of the study, analyzed and interpreted the data, and drafted and revised the manuscript. AJE mentored the design of the study, sought funding resources, and revised the manuscript.

Conflicts of Interest

The authors have no conflicts to report.

Multimedia Appendix 1
A. Entrance survey completed by all participants. B. Exit survey completed by all participants.
[DOCX File , 19 KB - formative_v6i3e28353_app1.docx]

Multimedia Appendix 2
Daily surveys delivered to all participants via text message during the control period (A), and intervention period (B).
[DOCX File , 19 KB - formative_v6i3e28353_app2.docx]

References

1. Liu Y, Wheaton AG, Chapman DP, Cunningham TJ, Lu H, Croft JB. Prevalence of healthy sleep duration among adults--United States, 2014. MMWR Morb Mortal Wkly Rep 2016 Feb 19;65(6):137-141 [FREE Full text] [doi: 10.15585/mmwr.mm6506a1] [Medline: 26890214]

2. Shockey TM, Wheaton AG. Short sleep duration by occupation group - 29 States, 2013-2014. MMWR Morb Mortal Wkly Rep 2017 Mar 03;66(8):207-213 [FREE Full text] [doi: 10.15585/mmwr.mm6608a2] [Medline: 28253230]

3. Di Milia L, Waage S, Pallesen S, Bjorvatn B. Shift work disorder in a random population sample--prevalence and comorbidities. PLoS One 2013;8(1):e55306 [FREE Full text] [doi: 10.1371/journal.pone.0055306] [Medline: 23372847]

4. Ganesan S, Magee M, Stone JE, Mulhall MD, Collins A, Howard ME, et al. The impact of shift work on sleep, alertness and performance in healthcare workers. Sci Rep 2019 Mar 15;9(1):4635 [FREE Full text] [doi: 10.1038/s41598-019-40914-x] [Medline: 30874565]

5. Kerkhof GA. Shift work and sleep disorder comorbidity tend to go hand in hand. Chronobiol Int 2018 Feb;35(2):219-228. [doi: 10.1080/07420528.2017.1392552] [Medline: 29157012]

6. Wickwire EM, Geiger-Brown J, Scharf SM, Drake CL. Shift work and shift work sleep disorder: clinical and organizational perspectives. Chest 2017 May;151(5):1156-1172 [FREE Full text] [doi: 10.1016/j.chest.2016.12.007] [Medline: 28012806]

7. Job Flexibilities and Work Schedules Summary. US Bureau of Labor Statistics. 2019 Sep 24. URL: https://www.bls.gov/news.release/flex2.nr0.htm [accessed 2020-09-20]

8. Basner M, Dinges DF, Shea JA, Small DS, Zhu J, Norton L, et al. Sleep and alertness in medical interns and residents: an observational study on the role of extended shifts. Sleep 2017 Apr 01;40(4):zsx027 [FREE Full text] [doi: 10.1093/sleep/zsx027] [Medline: 28329124]

9. Arzalier-Daret S, Buléon C, Bocca M, Denise P, Gérard JL, Hanouz J. Effect of sleep deprivation after a night shift duty on simulated crisis management by residents in anaesthesia. A randomised crossover study. Anaesth Crit Care Pain Med 2018 Apr;37(2):161-166. [doi: 10.1016/j.accpm.2017.05.010] [Medline: 28882740]

10. Kalmbach DA, Arnedt JT, Song PX, Guille C, Sen S. Sleep disturbance and short sleep as risk factors for depression and perceived medical errors in first-year residents. Sleep 2017 Mar 01;40(3):zsw073 [FREE Full text] [doi: 10.1093/sleep/zsw073] [Medline: 28369654]

11. Trockel MT, Menon NK, Rowe SG, Stewart MT, Smith R, Lu M, et al. Assessment of physician sleep and wellness, burnout, and clinically significant medical errors. JAMA Netw Open 2020 Dec 01;3(12):e2028111 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.28111] [Medline: 33284339]

12. Robbins R, Weaver M, Quan S, Rosenberg E, Barger L, Czeisler C, et al. Employee sleep enhancement and fatigue reduction programs: analysis of the 2017 CDC Workplace Health in America poll. Am J Health Promot 2021 May;35(4):503-513. [doi: 10.1177/0890117120969091] [Medline: 33172286]

13. Basner M, Dinges D. Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. Sleep 2011 May 01;34(5):581-591 [FREE Full text] [doi: 10.1093/sleep/34.5.581] [Medline: 21532951]

14. Trockel M, Bohman B, Lesure E, Hamidi MS, Welle D, Roberts L, et al. A brief instrument to assess both burnout and professional fulfillment in physicians: reliability and validity, including correlation with self-reported medical errors, in a sample of resident and practicing physicians. Acad Psychiatry 2018 Feb;42(1):11-24 [FREE Full text] [doi: 10.1007/s40596-017-0849-3] [Medline: 29196982]

15. Kalmbach DA, Fang Y, Arnedt JT, Cochran AL, Deldin PJ, Kaplin AI, et al. Effects of sleep, physical activity, and shift work on daily mood: a prospective mobile monitoring study of medical interns. J Gen Intern Med 2018 Jun;33(6):914-920 [FREE Full text] [doi: 10.1007/s11606-018-4373-2] [Medline: 29542006]

16. Peterson SA, Wolkow AP, Lockley SW, O'Brien CS, Qadri S, Sullivan JP, et al. Associations between shift work characteristics, shift work schedules, sleep and burnout in North American police officers: a cross-sectional study. BMJ Open 2019 Dec 01;9(11):e030302 [FREE Full text] [doi: 10.1136/bmjopen-2019-030302] [Medline: 31791964]

17. Stewart NH, Arora VM. The impact of sleep and circadian disorders on physician burnout. Chest 2019 Nov;156(5):1022-1030 [FREE Full text] [doi: 10.1016/j.chest.2019.07.008] [Medline: 31352036]

18. Weaver MD, Robbins R, Quan SF, O'Brien CS, Viyaran NC, Czeisler CA, et al. Association of sleep disorders with physician burnout. JAMA Netw Open 2020 Oct 01;3(10):e2023256 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.23256] [Medline: 33125492]

19. Circadian Adaptation to Shift Work. American Academy of Sleep Medicine. URL: https://j2vjt3dnbra3ps7ll1clb4q2-wpengine. netdna-ssl.com/wp-content/uploads/2019/02/ProviderFS_ShiftChanges_19.pdf [accessed 2022-02-11]

20. Schlafer O, Wenzel V, Högl B. [Sleep disorders among physicians on shift work]. Anaesthesist 2014 Nov;63(11):844-851. [doi: 10.1007/s00101-014-2374-z] [Medline: 25213642]

21. Kecklund G, Axelsson J. Health consequences of shift work and insufficient sleep. BMJ 2016 Nov 01;355:i5210. [doi: 10.1136/bmj.i5210] [Medline: 27803010]

22. Patterson PD, Mountz KA, Budd CT, Bubb JL, Hsin AU, Weaver MD, et al. Impact of shift work on blood pressure among emergency medical services clinicians and related shift workers: A systematic review and meta-analysis. Sleep Health 2020 Jun;6(3):387-398. [doi: 10.1016/j.sleh.2020.03.006] [Medline: 32354630]

23. Vetter C, Devore EE, Wegrzyn LR, Massa J, Speizer FE, Kawachi I, et al. Association between rotating night shift work and risk of coronary heart disease among women. JAMA 2016 Apr 26;315(16):1726-1734 [FREE Full text] [doi: 10.1001/jama.2016.4454] [Medline: 27115377]

24. Yau A, Haque M. Shiftwork association with cardiovascular diseases and cancers among healthcare workers: a literature review. Medeni Med J 2019;34(4):387-395 [FREE Full text] [doi: 10.5222/MMJ.2019.54775] [Medline: 32821466]

25. Francis MN, Wishart IM, Williamson T, Iverach R. Use of pharmacologic sleep aids and stimulants among emergency medicine staff physicians in a Canadian tertiary care setting: a web-based survey. Ann Emerg Med 2019 Apr;73(4):325-329. [doi: 10.1016/j.annemergmed.2018.07.006] [Medline: 30146443]

26. Shy BD, Portelli I, Nelson LS. Emergency medicine residents' use of psychostimulants and sedatives to aid in shift work. Am J Emerg Med 2011 Nov;29(9):1034-6.e1. [doi: 10.1016/j.ajem.2010.06.004] [Medline: 20708878]

27. Sadeghniiat-Haghighi K, Bahrami H, Aminian O, Meysami A, Khajeh-Mehrizi A. Melatonin therapy in shift workers with difficulty falling asleep: A randomized, double-blind, placebo-controlled crossover field study. Work 2016 Sep 27;55(1):225-230. [doi: 10.3233/WOR-162376] [Medline: 27612057]

28. Jorgensen KM, Witting MD. Does exogenous melatonin improve day sleep or night alertness in emergency physicians working night shifts? Ann Emerg Med 1998 Jun;31(6):699-704. [doi: 10.1016/s0196-0644(98)70227-6] [Medline: 9624308]

29. Dorrian J, Heath G, Sargent C, Banks S, Coates A. Alcohol use in shiftworkers. Accid Anal Prev 2017 Feb;99(Pt B):395-400. [doi: 10.1016/j.aap.2015.11.011] [Medline: 26621201]

30. Richter K, Peter L, Rodenbeck A, Weess HG, Riedel-Heller SG, Hillemacher T. Shiftwork and alcohol consumption: a systematic review of the literature. Eur Addict Res 2021;27(1):9-15 [FREE Full text] [doi: 10.1159/000507573] [Medline: 32454482]

31. Keller J, Kwasnicka D, Klaiber P, Sichert L, Lally P, Fleig L. Habit formation following routine-based versus time-based cue planning: A randomized controlled trial. Br J Health Psychol 2021 Sep;26(3):807-824. [doi: 10.1111/bjhp.12504] [Medline: 33405284]

32. Robbins R, Krebs P, Rapoport DM, Jean-Louis G, Duncan DT. Examining use of mobile phones for sleep tracking among a national sample in the USA. Health Commun 2019 May;34(5):545-551. [doi: 10.1080/10410236.2017.1422104] [Medline: 29334765]

33. Robbins R, Seixas A, Masters LW, Chanko N, Diaby F, Vieira D, et al. Sleep tracking: A systematic review of the research using commercially available technology. Curr Sleep Med Rep 2019;5(3):156-163 [FREE Full text] [doi: 10.1007/s40675-019-00150-1] [Medline: 33134038]

34. Espie CA, Emsley R, Kyle SD, Gordon C, Drake CL, Siriwardena AN, et al. Effect of digital cognitive behavioral therapy for insomnia on health, psychological well-being, and sleep-related quality of life: a randomized clinical trial. JAMA Psychiatry 2019 Jan 01;76(1):21-30 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.2745] [Medline: 30264137]

35.   Morin CM. Profile of somryst prescription digital therapeutic for chronic insomnia: overview of safety and efficacy. Expert Rev Med Devices 2020 Dec;17(12):1239-1248. [doi: 10.1080/17434440.2020.1852929] [Medline: 33226269]
36.   Riedy SM, Smith MG, Rocha S, Basner M. Noise as a sleep aid: A systematic review. Sleep Med Rev 2021 Feb;55:101385. [doi: 10.1016/j.smrv.2020.101385] [Medline: 33007706]
37.   Montgomery-Downs HE, Insana SP, Miller EA. Effects of two types of ambient sound during sleep. Behav Sleep Med 2010;8(1):40-47. [doi: 10.1080/15402000903425686] [Medline: 20043248]
38.   Sweity S, Finlay A, Lees C, Monk A, Sherpa T, Wade D. SleepSure: a pilot randomized-controlled trial to assess the effects of eye masks and earplugs on the quality of sleep for patients in hospital. Clin Rehabil 2019 Feb;33(2):253-261. [doi: 10.1177/0269215518806041] [Medline: 30322272]
39.   Hansen IP, Langhorn L, Dreyer P. Effects of music during daytime rest in the intensive care unit. Nurs Crit Care 2018 Jul;23(4):207-213. [doi: 10.1111/nicc.12324] [Medline: 29159864]

Abbreviations

ED: emergency department
EM: emergency medicine
IRB: institutional review board
PVT: psychomotor vigilance test
SWSD: shiftwork sleep disorder

XSL•FO
RenderX

<u>Original Paper</u>

Reversing the Antibiotic Resistance "Yelp Effect" Through the Use of Emotionally Framed Responses to Negative Reviews of Providers: Questionnaire Study

Monique Mitchell Turner[1], PhD; Hyesun Choung[1], PhD; Quoc-Ha Hannah Mai Bui[2], MPH; Paige Beck[2], MPH; Hera Ashraf[2], MPH

[1]College of Communication Arts and Sciences, Michigan State University, East Lansing, MI, United States
[2]Milken Institute School of Public Health, George Washington University, Washington, DC, United States

Corresponding Author:
Quoc-Ha Hannah Mai Bui, MPH
Milken Institute School of Public Health
George Washington University
950 New Hampshire Ave NW
Washington, DC, 20052
United States
Phone: 1 202 994 7400
Email: hmbui@gwu.edu

Abstract

Background: The overuse of antibiotics has rapidly made antimicrobial resistance a global public health challenge. There is an emerging trend where providers who perceive that their patients expect antibiotics are more likely to prescribe antibiotics unprompted or upon request. Particularly, health care providers have expressed concern that dissatisfied patients will provide disparaging online reviews, therefore threatening the reputation of the practice. To better deal with the negative reviews and inform patients, some health care staff directly respond to patients' online feedback. Engaging with patients' online reviews gives providers an opportunity to prevent reputational damage and improve patients' understanding of the antibiotic resistance problem.

Objective: We aim to test the effectiveness of different response strategies to the negative patient online reviews on the readers' perceptions of the health care provider and their perceptions related to antibiotics resistance.

Methods: Two experiments were conducted to examine the impact of message tactics (apologizing, inducing fear or guilt) that can be employed by health care providers when responding to patients' negative online feedback related to not receiving an antibiotic.

Results: Overall, our results demonstrated positive impacts of responding to patients' online reviews. In study 1, we found apologetic messaging and use of emotional appeals in the response were effective in making readers feel more favorable toward the message. Readers also expressed a greater credibility perception toward the provider and willingness to visit the clinic when emotional appeals were used. Findings from study 2 largely supported the effectiveness of a fear-based response in improving the readers' credibility perceptions and willingness to visit the clinic. The fear-inducing information was particularly effective among parent readers.

Conclusions: This paper demonstrated that a strategic response to online patient complaints could prevent reputational damage and minimize the potential negative impacts of the review. The results also glean insight into the step toward developing a novel intervention—crafting a persuasive response to patients' negative feedback that can help improve the understanding of antibiotic resistance problems.

KEYWORDS

Introduction

There are at least 2.8 million antibiotic-resistant infections per year in the United States, and as many as 35,000 people die each year due to such infections [1]. Nearly 1 in 3 antibiotics prescribed at outpatient facilities is found to be unnecessary [2]. Reducing the overuse of antibiotics in such facilities, like urgent care centers, can help to lower risks associated with antimicrobial resistance. While the judicious use of antibiotics has been widely promoted, there are still barriers to antibiotic stewardship practices [3]. The most commonly cited reason for unnecessary prescribing is providers' perception of patient expectation and satisfaction [4,5]. Studies have consistently reported that patient demand greatly influences antibiotic prescribing decisions [6-8]. Likewise, a major determinant of prescribing antibiotics in pediatric care is parents' expectations and pressure [9,10]. Improving communication between health care providers and patients is suggested as one of the ways to reduce overprescribing [11,12].

This perception of pressure from patients can be exacerbated by other external factors. For example, as patients are increasingly behaving as consumers by posting their experiences with health care providers to websites such as Yelp.com [13], providers may fear that their patient will post a disparaging online review of the clinic and provider when requested antibiotics were not provided. In fact, these felt pressures have become so well known that Wired Magazine termed it "the Yelp effect" [14]. A study showed that 65% of Americans are aware of these online ratings and 37% had avoided a physician based on negative reviews [15]. Anxious over potential negative online reviews, providers might be pressured to comply with unrealistic patient expectations such as prescribing antibiotics even when not warranted.

For the clinics of providers, it is important to engage with patient feedback as those reviews can influence future patients' visits [16]. For example, if the complaint is about not receiving requested antibiotics during the visit, providers can think of it as an opportunity to improve the patient experience and inform the patient about the potential consequences of taking unnecessary treatments or antibiotics. Nonetheless, few strategies have been tested to assess how they might positively impact patients' antibiotic perceptions and expectations and ultimately act as a step in stemming the tide of antibiotic resistance. In fact, there are no studies of this ilk in the domain of the so-called Yelp effect, yet there are effective risk and health communication strategies that should be tested in this domain. Scholars have long proposed that in order to garner attention, increase message scrutiny, and change attitudes, emotional appeals can be useful [17,18]. Two emotional appeals that have received attention are fear and guilt.

Fear appeals, which communicate an impending threat coupled with behavioral recommendations to mitigate the threat, have consistently been shown to effectively change attitudes [19]. If the fear appeal effectively communicates the threat and receivers believe the response is an effective measure they can take, then message receivers judge that the risk is dangerous and can affect their lives. A meta-analysis investigated fear appeal's effectiveness for influencing attitudes, intentions, or behaviors and found that fear appeals are effective [20]. In addition, practitioners find that fear appeals can be effectively used to persuade audiences when used appropriately [21]. When communicating antibiotic resistance, fear can be aroused by emphasizing the potential harm or danger that will befall individuals if they take unnecessary antibiotics. In this case, we argue that fear messaging can increase the perception of antibiotic resistance risk. Thus, the reader of a negative review would believe that a provider who did not provide antibiotics was being careful and protecting the patient. This could mitigate unfavorable effects of the negative review.

> *H1: A response from the clinic that features fear-inducing antibiotic resistance information will have a positive impact on (a) message favorability, (b) provider credibility, and (c) increased willingness to visit the clinic in the future relative to no response or a simple apology (ie, apologetic response).*

Guilt appeals, though, work through a different mechanism. Guilt appeals communicate that a harm could be potentially caused by our own actions or inactions and are commonly used in health and risk communication [22-25]. These subtle messages remind audiences of their moral code (eg, do not harm others) and can spark both attitude change and behavior change. Turner's research [22,26,27] revealed that when messages convey a discrepancy between individuals' moral norms and their potential future behavior, they are less likely to engage in the harmful behavior. In this case, people would be reminded that pressuring doctors for antibiotics ultimately affects the antibiotic health crisis. Guilt is seen as a prototypical moral emotion, and it is repeatedly used in persuasion campaigns for its behavioral consequences [24].

> *H2: A response from the clinic that features guilt-inducing antibiotic resistance information will have a positive impact on (a) making readers feel favorable toward the message, (b) increasing the credibility of the reviewed provider, and (c) increase the willingness to visit the clinic in the future relative to no response or an apologetic response.*

As indicated in hypotheses 1 and 2, one type of baseline comparison group we test is a simple apology. Apologies are considered to be the essential response component that service providers can use to handle consumer complaints [28,29]. Health care providers can engage in complaints by apologizing and taking responsibility for patients' negative experiences. This can help them to reclaim a favorable impression as prior research has documented that apologies are effective for reestablishing trust [30,31]. Accordingly, we hypothesized the following:

> *H3: A response from the clinic apologizing for patients' negative experience will have a positive impact on (a) making readers feel favorable toward the message, (b) increasing the credibility of the reviewed provider, and (c) increasing the willingness to visit the clinic in the future relative to providing no response to the review.*

Study 1

Introduction

Based on the vast literature on emotionally framed health and risk communication, we tested strategic communication regarding antibiotic resistance in the form of replies to negative patient online reviews. We test the effectiveness of fear, guilt, and apologetic message strategies used when responding to a patient's negative online reviews complaining about not receiving requested antibiotics during its visit to the local urgent care. We accomplish these objectives by conducting a randomized controlled experiment with the goal of examining the impact of different response message strategies on readers' perceptions of the health care providers and perceptions and knowledge about antibiotics resistance.

Methods

A between-subjects experimental design with 4 message conditions—control (no response), apology messaging, fear-based messaging, and guilt-based messaging—was used. Participants were asked to imagine that they have moved to a new town and need to find a local urgent care. Then, participants viewed a review site (like Yelp, but deidentified) where a patient provides a negative review of an urgent care. The patient arrived at urgent care with a cold and a sore throat but did not receive an antibiotic. The wording of the fictitious online review was based on real online reviews. Participants also saw the urgent care clinic representative's (staff or care provider) response to the negative review based on the experimental condition. See Multimedia Appendix 1 for stimulus. Participants were allowed to view this page as long as they needed to. Afterward, participants were asked to complete a short survey and asked their favorability toward the response; the credibility of the clinic, provider and staff; and willingness to visit the clinic in the future (see Multimedia Appendix 2). Participants were recruited through the Amazon Mechanical Turk platform. A total of 216 adults (55% (119/216) male, 81% (175/216) White) from around the nation were paid to take part in a 15-minute experiment. About 64% (138/216) of the participants were aged 30 to 49 years, and 42% (91/216) had a Bachelor's degree.

Results

To test the effect of response strategies on outcome variables, we ran a series of 1-way analyses of variance (ANOVAs) with Tukey post hoc testing. There was a statistically significant difference in readers' favorability toward the message ($F_{3, 212}=29.81$, $P<.001$), perceived provider credibility ($F_{3, 212}=3.66$, $P=.01$), and willingness to visit the reviewed clinic $F_{3, 212}=4.01$, $P=.008$) across message conditions (see Table 1).

Table 1. Effects of response strategies by messaging condition in study 1 (Note: Means in a row sharing superscripts are significantly different from one another).

Variable	Control (1)		Apology (2)		Fear (3)		Guilt (4)		η^2
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Favorability toward the message	2.90[*]	0.78	3.86[#]	0.84	4.44[^]	0.89	4.30[^]	1.00	.297
Provider credibility perception	3.65[*]	1.17	3.71[*]	1.08	4.20[#]	1.06	4.16[*,#]	1.11	.049
Willingness to visit the clinic	3.56[*,#]	0.98	3.36[*,#]	0.98	3.89[*,^]	1.19	3.98[*,^]	1.08	.054

These data support the hypothesis that a fear-inducing message was effective in making readers feel more favorable toward the response (mean 4.44 [SD .89]) compared to the control group (P<.001). This response was also more effective than an apology message (mean 3.86 [SD .84], P=.003). H1a was supported. Furthermore, the fear message increased readers' credibility perception toward the reviewed provider (mean 4.20 [SD 1.06]) compared to the control group (mean 3.65 [SD 1.17], P=.06), and the apology message condition (mean 3.71 [SD 1.08], P=.09). H1b was supported. Respondents' willingness to visit the clinic in the future was only significantly different between the fear (mean 3.89 [SD 1.19]) condition and the apology condition (mean 3.36 [SD .98], P=.04). The mean difference between the fear-inducing condition and the control condition was not significant. Accordingly, H1c was partially supported.

Similarly, guilt-inducing messaging made readers feel more favorable toward the response (mean 4.30 [SD 1.00]) relative to apologetic messaging (mean 3.86 [SD .84], P=.047). Thus, H2a was supported. The guilt message did not increase readers' credibility perception toward the provider (H2b not supported). When a response offered guilt-based antibiotic resistant information, readers were more willing to visit the clinic in the future (mean 3.98 [SD 1.08]) compared to the apology message condition (mean 3.36 [SD .98], P=.01). Nonetheless, a statistically significant difference was not observed when the guilt condition was compared to the control condition (mean 3.56 [SD .98]). Therefore, H2c was partially supported.

We found that the readers who read an apology message (mean 3.86 [SD .84]) from a clinic, which received a negative review from a past patient, felt more favorable toward the response compared to the readers who read the negative patient review with no reply from the clinic (mean 2.90 [SD .78], P<.001). However, there was no significant effect of apology messaging on the credibility perception and willingness to visit the clinic. Thus, H3a was supported, but H3b and H3c were not supported.

The findings from study 1 provide empirical support for the idea that engaging with patients' negative reviews with an effective and evidence-based communication strategy can prevent potential reputational damage and improve the patient-provider relationship. In doing so, our findings suggest that rather than just apologizing for patients' negative experiences (which implies admitting blame), using emotional appeals would further enhance the effectiveness of the message

in building favorable attitudes, credibility, and increasing potential patients' chance to visit the clinic in the future.

Specifically, we found that apologizing for a patient's negative experience can make readers feel more favorable, but it did not increase credibility perceptions or willingness to visit the clinic. If the goal of the communication is to restore credibility and increase patient visits, emotion-based messaging is more effective than not responding or apologizing. We found the fear-inducing information related to antibiotic resistance was particularly effective in increasing the credibility perception of the provider.

These findings can help empower managers, owners, and providers within urgent care clinics to deal with the pressures patients might exert when they want antibiotics. It is critical that providers feel confident that they can engage in clinically sound practices without harming their, or the clinic's, credibility. It is critical to assess the replicability of these findings, though, so we replicated and extended this study with a second national panel of participants.

Study 2

Introduction

A second study was conducted to test the robustness of some of the key findings from the first study. In a pediatric care setting, parent pressure to prescribe antibiotics for their child is a barrier for working toward a strict adherence of antibiotic prescription [9,10]. Thus, in the second experiment, we test the effectiveness of response strategies when a negative online review is left by a parent who did not get antibiotics prescribed for their sick child. In study 2, we also compare which type of message strategy (control vs apology vs fear vs guilt) is more effective in altering readers' antibiotics expectations and perceptions.

In study 1, we examined if strategic messaging could improve readers' perceptions toward the reviewed provider. To test the robustness of the findings, we suggest the following hypotheses:

> *H1: A response from the clinic that features fear-inducing antibiotic resistance information will increase (a) the credibility of the reviewed provider and (b) the willingness to visit the clinic in the future relative to providing no response or an apologetic response.*
>
> *H2: A response from the clinic that features guilt-inducing antibiotic resistance information will increase (a) the credibility of the reviewed provider and (b) the willingness to visit the clinic in the future relative to providing no response or an apologetic response.*
>
> *H3: A response from the clinic apologizing for patients' negative experience will increase (a) the credibility of the reviewed provider and (b) the willingness to visit the clinic in the future relative to providing no response to the review.*

In study 2, we also test if the clinic's response can be an effective educational intervention that improves people's understanding of the antibiotic resistance problem. We predict that an effective message will lower people's unrealistic expectations for antibiotics and reduce the conception of better to be safe than sorry (BSTS). The latter, BSTS, is the common misperception that even in the cases when taking antibiotics are unlikely to be effective; patients perceive that there is some chance that antibiotics might be effective and have little risk. Patients prefer to take antibiotics since doing so provides the possibility of getting better, and it is often the patients' rationale for requesting antibiotics [32]. Thus, we try to answer the following question:

> *RQ1: Which response message strategy (apology vs fear-inducing vs guilt-inducing messaging) is more effective in (a) lowering people's antibiotics expectations and (b) reducing BSTS misconception compared to the control group condition?*

As the online review used in study 2 features a parent's experience with a clinic, we also evaluate whether there is any difference in the effectiveness of the response strategies between parent readers and nonparent readers with the following research questions.

> *RQ2: Is there any difference in the level of (a) credibility perception, (b) willingness to visit, (c) antibiotics expectations, and d) BSTS misconception between parent readers and nonparent readers?*
>
> *RQ3: Is there any difference in the effects of response strategies on the (a) credibility perception, (b) willingness to visit, (c) antibiotics expectations, and (d) BSTS misconception while controlling for the parental status?*

Methods

Study Overview

A 4 (message type: control, apology, fear-based, guilt-based) × 2 (parental status: parent vs nonparent) between-subjects experimental design was used. Multimedia Appendix 1 includes the stimulus used in the study. Participants aged 18 years and older and living in the United States were recruited. As with study 1, we used the MTurk survey platform for this 10-minute experimental survey. A total of 400 US adults took the survey; after removing incomplete responses, a total of 390 responses were used in the analysis. The median age group of participants was 30 to 49 years, and 62% (242/390) of participants were male. About 58% (226/390) of participants identified their race as White and 58% (226/390) had a Bachelor's degree. A summary of survey measurements is presented in Multimedia Appendix 2.

Data Analysis

A 2-way ANOVA was conducted to determine the effects of message strategies and parental status for each outcome variable. Table 2 reports a summary of test results. An additional simple effect analysis was conducted to probe interaction effects.

XSL·FO

RenderX

Table 2. Summary table for 2-way analysis of variance of the effects of messaging strategies and parental status in study 2.

Variable and source	MS[a]	F score	P value	η_p^2
Provider credibility perception				
Message[b]	5.96	4.02	.008	.031
Parental status[c]	33.24	22.45	<.001	.056
Message × parental status[b]	3.19	2.15	.093	.017
Willingness to visit the clinic				
Message[b]	8.23	19.36	<.001	.033
Parental status[c]	36.57	19.36	<.001	.048
Message × parental status[b]	7.09	3.75	.011	.029
Antibiotic expectation				
Message[b]	0.49	0.56	.640	.004
Parental status[c]	2.34	2.7	.102	.007
Message × parental status[b]	0.625	0.72	.541	.006
BSTS[d] misconception				
Message[b]	1.13	1.28	.281	.010
Parental status[c]	24.66	27.99	<.001	.068
Message × parental status[b]	0.606	0.69	.560	.005

[a]MS: mean squares.

[b]$df = 3, 382$.

[c]$df = 1, 382$.

[d]BSTS: better safe than sorry.

Results

To test H1-H3, we first compared the mean scores of credibility perception and willingness to visit the clinic across the 4 message conditions (Table 3). We found that response strategies have statistically significant effects on readers' credibility perception of the provider and willingness to visit the clinic (see Table 2).

Table 3. Effects of response strategies by messaging condition in study 2 (Note: Means in a row sharing superscripts are significantly different from one another).

Variable	Control (1)		Apology (2)		Fear (3)		Guilt (4)		η^2
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Provider credibility perception	2.51[*]	1.13	2.92[*,#]	1.27	3.09[#,^]	1.29	2.82[*,#,^]	1.3	.028
Willingness to visit the clinic	2.61[*]	1.30	2.94[*,#]	1.48	3.26[#,^]	1.42	3.05[*,#,^]	1.5	.027
Antibiotic expectation	3.24[*]	0.94	3.25[*]	0.90	3.11[*]	0.92	3.15[*]	0.97	.004
BSTS misconception	2.91[*]	0.95	2.71[*]	0.96	2.76[*]	0.95	2.71[*]	1.01	.008

These data support the hypothesis that fear-inducing information can increase providers' credibility perception (mean 3.09 [SD 1.29], P=.006) and willingness to visit the clinic (mean 3.26 [SD 1.42], P=.008) when compared to the control group (credibility: mean 2.51 [SD 1.13]; willingness to visit: mean 2.61 [SD 1.30]), but the differences were not statistically significant compared to the apology message condition (mean 2.98 [SD1.27]). Thus, H1 was partially supported. The guilt-inducing response did not increase credibility perception.

Study 1 showed that the guilt message can increase readers' willingness to visit when compared to the apology condition. However, in study 2, which is in the pediatric context, the effect was not statistically significant. Accordingly, H2 was not supported. Regarding H3, there were no statistically significant effects of the apology message on the credibility perception or willingness to visit the clinic (H3 not supported), repeating the findings from study 1.

XSL•FO
RenderX

To answer RQ1, we compared the mean scores of the expectations for antibiotics and BSTS misconception (see Table 3) across the message conditions. We found no significant main effects of message strategies on antibiotic expectation and BSTS misconception.

Two participant groups (parents vs nonparents) were compared to answer RQ2 (see Table 4). Participants' parental status had a statistically significant impact on provider credibility perception, willingness to visit the clinic, and the BSTS misconception. Parents perceived the provider less credible (mean 2.56 [SD 1.19], P<.001) compared to nonparents (mean 3.13 [SD 1.29]). Parents were also less willing to visit the clinic in the future (mean 2.69 [SD1.42], P<.001) than nonparents (mean 3.26 [SD 1.40]). Parent readers exhibited greater BSTS misconception (mean 3.01 [SD .86], P<.001) than nonparent participants (mean 2.52 [SD 1.01]). There was no statistically

significant difference observed in their expectations of getting antibiotics.

Finally, a statistically significant interaction effect between parental status and the message strategy was detected on individuals' willingness to visit the clinic in the future (RQ3, see Figure 1). For the parent participants ($F_{3, 384}=3.51$, P=.02), the response that elicited fear increased their willingness to visit the clinic (mean 3.18 [SD .20]) significantly more than the control (mean 2.52 [SD .20], P=.02) and the apology (mean 2.31 [SD .20], P=.002) message conditions. For the nonparent participants ($F_{3, 384}=3.93$, P=.009), all 3 types of response messages increased their willingness to visit the clinic (apology: mean 3.56 [SD .20], P=.002; fear: mean 3.33 [SD .19], P=.02; guilt: mean 3.49 [SD .21], P=.006) when compared to the control group condition (mean 2.69 [SD .20]). There was no statistically significant difference among the three message conditions for nonparents.

Table 4. Mean and standard deviations for effects of parental status in study 2.

Variable	Parent		Nonparent		η^2	P value
	Mean	SD	Mean	SD		
Provider credibility perception	2.56	1.18	3.13	1.29	.051	<.001
Willingness to visit the clinic	2.69	1.42	3.26	1.40	.040	<.001
Antibiotic expectation	3.26	0.97	3.11	0.88	.007	.11
BSTS[a] misconception	3.01	0.86	2.52	1.01	.065	<.001

[a]BSTS: better safe than sorry.

Figure 1. Interaction effects between message condition and parental status on willingness to visit the clinic in Study 2.
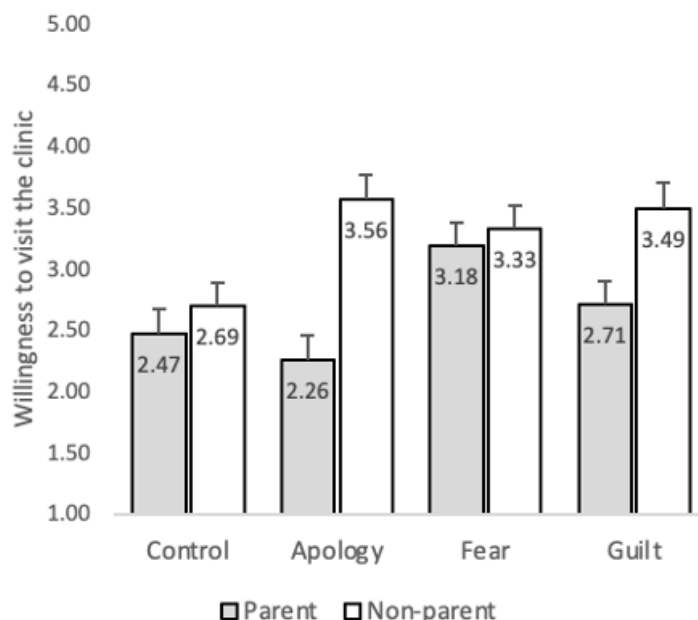
Discussion

Principal Findings

Emotional appeals can be used to better communicate the information about the risks and consequences of inappropriate

use of antibiotics. Appealing to human emotion has long been known to help garner attention, engage message receivers cognitively, and change beliefs and intentions [18]. This study showed that such emotion-based tactics need not be solely used in public service announcements; rather, such known message strategies can also be used in responses to online reviews.

Providers and clinic owners can and should communicate good clinical practice whenever they can do so. Instead of being worried about negative online reviews, it is possible that these moments can be leveraged as teachable moments, and we believe this paper provides insight into how to develop evidence-based online interventions [33,34].

Study 1 was conducted in the primary care context. This study revealed that fear-based messaging can increase message favorability, provider credibility, and willingness to go to the clinic compared to a simple apology. Given that the apologetic response is the most common response to negative online reviews, this information is critical for clinics. These results can be explained by the studies that showed that when communication of the severity of a health threat is combined with communication of susceptibility to the threat, message receivers are more likely to engage with the message content [17]. Particularly if the message communicates a solution (ie, efficacy), message receivers will engage in a process known as danger control whereby they are fearful of the health threat, understand the threat can be averted, and want to engage in behaviors that would control and prevent it [35].

The findings from study 2 further demonstrated the effectiveness of using a fear-frame when responding to a patient's negative online review. Another important finding was that while different message strategies were equally effective in increasing nonparents' willingness to visit, only the fear-inducing message had an impact on parent readers. Thus, when responding to a pediatrics online review often left and read by parents, fear messaging, demarked by communicating susceptibility to, and severity of a risk, can help to mitigate potential negative consequences of a negative online review.

Moreover, we found that parents were overall more influenced by the negative online review than nonparents. Parents perceived the provider as less credible, and they were less willing to visit the clinic than nonparents. This could be because parents are generally concerned about doing the right thing for their child. When it comes to medical decisions, they experience greater uncertainty and anxiety with their decisions for their ill child [36]. Parents held greater BSTS misperception (taking antibiotics has little risk associated and increases the chance of getting better even when it might not be needed) because they are more worried about missing an opportunity to treat their child's serious illness than a long-term side effects of antibiotics.

Finally, the emotion-based response strategies did not have statistically significant effects on lowering individuals' antibiotic expectations or correcting the BSTS misperception. Although we found limited educational effects of the response messages, as changing patients' perception is critical to reduce the overprescribing of antibiotics, future studies should examine the effects of different message strategies and interventions that can improve people's understanding of antibiotic resistance.

To this point, it is critical that expertise in messaging strategies is sought before clinicians design responses to online reviews (or develop message strategies generally). It is well known that certain messaging strategies can fail when used with the wrong audience or in the wrong context. In this study, we found that while nonparent readers were receptive to various messaging strategies, only fear-inducing messaging was effective among parent readers. Previous studies reported that a persuasive appeal's effectiveness varies by the audience, and aligning message strategies with the recipient's personal traits and emotional status can increase the message's impact [37,38]. Likewise, certain messaging strategies can fail when used with the wrong audience or in the wrong context [39]. Thus, practitioners should critically analyze their target audience and develop the best messaging strategies accordingly.

Limitations

There are limitations to this research. Certainly, these studies did not use probability-based samples (although they did use random assignment), and thus, we should be cautious when generalizing the findings. Additionally, this study only examined one iteration of fear or guilt; but guilt and fear messages vary a great deal in intensity. Overly intense messages can cause anger and perceptions of being manipulated—so such strategies must be used subtly and thoughtfully. We were also mainly interested in the effects of messages in both studies, but the tested outcome variables (ie, favorability, credibility perception, and willingness to visit) are interrelated. For instance, it is possible that the favorability or credibility perception may predict readers' desire to visit the clinic. Thus, future studies could further our findings by investigating a causal model among the outcome variables engaging in mediation analysis. Finally, it may be that clinics are uncomfortable replying to negative online reviews at all. Although we hope that these data provide confidence that replying is a good idea, we recognize that some clinic managers will remain reticent to do so. In those cases, the emotional appeals here could be used to inform in-clinic messages that might change or adjust expectations prior to a negative review happening.

Conclusion

It is imperative that providers feel empowered to follow clinically sound practices. In this case, that means not feeling pressured to prescribe antibiotics when they are not needed. Health communication theories can provide strategies that can be used to assist in these efforts. For example, some clinics have tested the impacts of displaying posters using a commitment strategy [40], as a well-known strategy in communication and psychology, on changing patient expectations [41]. Overall, the strategy has shown promising results in affecting patient perceptions and expectations. Here, we examined the role of emotionally framed messages. Health communication scholars have tested emotional appeals on attitude change for nearly a century, finding that they can have a positive impact at how people perceive health risks. In this case, the key outcomes were to increase the credibility of a provider and willingness to visit a clinic. More studies of this ilk must be conducted going forward so that we can arm providers with efficacious communication devices that help them do their job.

Acknowledgments

Authors' Contributions

MMT, QHMB, PB, and HA collected the data. MMT, HC, and QHMB analyzed the data and wrote the first draft of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of stimuli.
[DOCX File , 525 KB - formative_v6i3e26122_app1.docx]

Multimedia Appendix 2
Summary survey questionnaires and reliability coefficients.
[DOCX File , 17 KB - formative_v6i3e26122_app2.docx]

References

1. Antibiotic resistance threatens everyone. Atlanta: Centers for Disease Control and Prevention URL: https://www.cdc.gov/drugresistance/index.html [accessed 2020-11-02]
2. 1 in 3 antibiotic prescriptions unnecessary. Atlanta: Centers for Disease Control and Prevention URL: https://www.cdc.gov/media/releases/2016/p0503-unnecessary-prescriptions.html [accessed 2020-11-02]
3. Fleming-Dutra KE, Hersh AL, Shapiro DJ, Bartoces M, Enns EA, File TM, et al. Prevalence of inappropriate antibiotic prescriptions among US ambulatory care visits, 2010-2011. JAMA 2016 May 03;315(17):1864-1873. [doi: 10.1001/jama.2016.4151] [Medline: 27139059]
4. Shapiro E. Injudicious antibiotic use: an unforeseen consequence of the emphasis on patient satisfaction? Clin Therapeutics 2002 Jan;24(1):197-204. [doi: 10.1016/s0149-2918(02)85015-9]
5. Sirota M, Round T, Samaranayaka S, Kostopoulou O. Expectations for antibiotics increase their prescribing: causal evidence about localized impact. Health Psychol 2017 Apr;36(4):402-409. [doi: 10.1037/hea0000456] [Medline: 28206788]
6. Macfarlane J, Holmes W, Macfarlane R, Britten N. Influence of patients' expectations on antibiotic management of acute lower respiratory tract illness in general practice: questionnaire study. BMJ 1997 Nov 08;315(7117):1211-1214 [FREE Full text] [doi: 10.1136/bmj.315.7117.1211] [Medline: 9393228]
7. Zetts RM, Stoesz A, Smith BA, Hyun DY. Outpatient antibiotic use and the need for increased antibiotic stewardship efforts. Pediatrics 2018 Jun 23;141(6):e20174124. [doi: 10.1542/peds.2017-4124] [Medline: 29793986]
8. Butler CC, Rollnick S, Pill R, Maggs-Rapport F, Stott N. Understanding the culture of prescribing: qualitative study of general practitioners' and patients' perceptions of antibiotics for sore throats. BMJ 1998 Sep 05;317(7159):637-642 [FREE Full text] [doi: 10.1136/bmj.317.7159.637] [Medline: 9727992]
9. Szymczak JE, Feemster KA, Zaoutis TE, Gerber JS. Pediatrician perceptions of an outpatient antimicrobial stewardship intervention. Infect Control Hosp Epidemiol 2014 Oct 10;35 Suppl 3(S3):S69-S78. [doi: 10.1086/677826] [Medline: 25222901]
10. Mangione-Smith R, McGlynn EA, Elliott MN, Krogstad P, Brook RH. The relationship between perceived parental expectations and pediatrician antimicrobial prescribing behavior. Pediatrics 1999 Apr 01;103(4 Pt 1):711-718. [doi: 10.1542/peds.103.4.711] [Medline: 10103291]
11. Poole NM. Judicious antibiotic prescribing in ambulatory pediatrics: communication is key. Curr Probl Pediatr Adolesc Health Care 2018 Nov;48(11):306-317. [doi: 10.1016/j.cppeds.2018.09.004] [Medline: 30389361]
12. Llor C, Bjerrum L. Antimicrobial resistance: risk associated with antibiotic overuse and initiatives to reduce the problem. Ther Adv Drug Saf 2014 Dec 16;5(6):229-241 [FREE Full text] [doi: 10.1177/2042098614554919] [Medline: 25436105]
13. Rastegar-Mojarad M, Ye Z, Wall D, Murali N, Lin S. Collecting and analyzing patient experiences of health care from social media. JMIR Res Protoc 2015 Jul 02;4(3):e78 [FREE Full text] [doi: 10.2196/resprot.3433] [Medline: 26137885]
14. McKenna M. The "Yelp Effect" of patient surveys is making us sick. Wired. URL: https://www.wired.com/story/the-yelping-of-the-american-doctor/ [accessed 2020-11-02]
15. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. JAMA 2014 Feb 19;311(7):734-735. [doi: 10.1001/jama.2013.283194] [Medline: 24549555]
16. Arendt F, Forrai M, Findl O. Dealing with negative reviews on physician-rating websites: an experimental test of how physicians can prevent reputational damage via effective response strategies. Soc Sci Med 2020 Dec;266:113422 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113422] [Medline: 33065499]

17. Allen W. Effect of fear-arousing communications. J Abnorm Psychol 1953 Jan;48(1):78-92. [doi: 10.1037/h0060732] [Medline: 13022199]

18. Turner M. Using emotional appeals in health messages. In: Cho H, editor. Health Communication Message Design: Theory and Practice. Thousand Oaks: SAGE Publications; 2011:59-71.

19. Dillard JP, Anderson JW. The role of fear in persuasion. Psychol Mark 2004 Nov;21(11):909-926. [doi: 10.1002/mar.20041]

20. Tannenbaum MB, Hepler J, Zimmerman RS, Saul L, Jacobs S, Wilson K, et al. Appealing to fear: a meta-analysis of fear appeal effectiveness and theories. Psychol Bull 2015 Nov;141(6):1178-1204. [doi: 10.1037/a0039729] [Medline: 26501228]

21. Simpson JK. Appeal to fear in health care: appropriate or inappropriate? Chiropr Man Therap 2017 Sep 20;25(1):27 [FREE Full text] [doi: 10.1186/s12998-017-0157-8] [Medline: 28932388]

22. Boudewyns V, Turner MM, Paquin RS. Shame-free guilt appeals: testing the emotional and cognitive effects of shame and guilt appeals. Psychol Mark 2013 Jul 23;30(9):811-825. [doi: 10.1002/mar.20647]

23. Becheur I, Valette-Florence P. The use of negative emotions in health communication messages: study of the effects of fear, guilt, and shame. Recherche et Applications en Marketing (English Edition) 2014 Oct 09;29(4):89-109. [doi: 10.1177/2051570714552620]

24. Graton A, Mailliez M. A theory of guilt appeals: a review showing the importance of investigating cognitive processes as mediators between emotion and behavior. Behav Sci (Basel) 2019 Nov 20;9(12):117 [FREE Full text] [doi: 10.3390/bs9120117] [Medline: 31756909]

25. Turner M. Using guilt to motivate individuals to adopt healthy habits. In: Oxford Research Encyclopedia of Communication. Oxford: Oxford University Press; 2017.

26. Turner MM, Mabry-Flynn A, Shen H, Jiang H, Boudewyns V, Payne D. The effects of guilt-appeal intensity on persuasive and emotional outcomes: the moderating role of sponsor motive. J Nonprofit Pub Sector Mark 2017 Jul 17;30(2):134-150. [doi: 10.1080/10495142.2017.1326345]

27. Turner MM, Underhill JC. Motivating emergency preparedness behaviors: the differential effects of guilt appeals and actually anticipating guilty feelings. Commun Q 2012 Sep;60(4):545-559. [doi: 10.1080/01463373.2012.705780]

28. Lee CH, Cranage DA. Toward understanding consumer processing of negative online word-of-mouth communication. J Hospitality Tourism Res 2012 Jul 03;38(3):330-360. [doi: 10.1177/1096348012451455]

29. Mitchell-Box K, Braun KL. Fathers' thoughts on breastfeeding and implications for a theory-based intervention. J Obstet Gynecol Neonatal Nurs 2012;41(6):E41-E50. [doi: 10.1111/j.1552-6909.2012.01399.x] [Medline: 22861175]

30. Bottom WP, Gibson K, Daniels SE, Murnighan JK. When talk is not cheap: substantive penance and expressions of intent in rebuilding cooperation. Org Sci 2002 Oct;13(5):497-513. [doi: 10.1287/orsc.13.5.497.7816]

31. Kim PH, Ferrin DL, Cooper CD, Dirks KT. Removing the shadow of suspicion: the effects of apology versus denial for repairing competence- versus integrity-based trust violations. J Appl Psychol 2004 Feb;89(1):104-118. [doi: 10.1037/0021-9010.89.1.104] [Medline: 14769123]

32. Broniatowski DA, Klein EY, May L, Martinez EM, Ware C, Reyna VF. Patients' and clinicians' perceptions of antibiotic prescribing for upper respiratory infections in the acute care setting. Med Decis Making 2018 May 30;38(5):547-561. [doi: 10.1177/0272989x18770664]

33. Turner M. The role of emotion in persuasion and risk communication. In: Doeveling K, von Scheve C, Konijin EA, editors. The Routledge Handbook of Emotions and Mass Media. London: Routledge; 2011:237-258.

34. Turner M. Emotional branding: what, when and why. In: Evans WD, editor. The Psychology of Branding. Hauppauge: Nova Science Publishers; 2013.

35. Witte K. Fear control and danger control: a test of the extended parallel process model (EPPM). Commun Mono 2009 Jun 02;61(2):113-134. [doi: 10.1080/03637759409376328]

36. Neill SJ, Jones CH, Lakhanpaul M, Roland DT, Thompson MJ. Parents' help-seeking behaviours during acute childhood illness at home. J Child Health Care 2014 Oct 08;20(1):77-86. [doi: 10.1177/1367493514551309] [Medline: 25296933]

37. Cesario J, Grant H, Higgins ET. Regulatory fit and persuasion: transfer from "Feeling Right.". J Pers Soc Psychol 2004 Mar;86(3):388-404. [doi: 10.1037/0022-3514.86.3.388] [Medline: 15008644]

38. Hirsh JB, Kang SK, Bodenhausen GV. Personalized persuasion: tailoring persuasive appeals to recipients' personality traits. Psychol Sci 2012 Jun 30;23(6):578-581. [doi: 10.1177/0956797611436349] [Medline: 22547658]

39. Bessarabova E, Turner MM, Fink EL, Blustein NB. Extending the theory of reactance to guilt appeals. Zeitschrift für Psychologie 2015 Oct;223(4):215-224. [doi: 10.1027/2151-2604/a000223]

40. Burger JM. The foot-in-the-door compliance procedure: a multiple-process analysis and review. Pers Soc Psychol Rev 1999 Dec 21;3(4):303-325. [doi: 10.1207/s15327957pspr0304_2] [Medline: 15661679]

41. Meeker D, Knight TK, Friedberg MW, Linder JA, Goldstein NJ, Fox CR, et al. Nudging guideline-concordant antibiotic prescribing: a randomized clinical trial. JAMA Intern Med 2014 Mar 01;174(3):425-431 [FREE Full text] [doi: 10.1001/jamainternmed.2013.14191] [Medline: 24474434]

Abbreviations

ANOVA: analysis of variance

BSTS: better to be safe than sorry

XSL·FO

RenderX

<u>Original Paper</u>

Improving Access to Behavioral Strategies to Improve Mental Well-being With an Entertaining Breakfast Show App: Feasibility Evaluation Study

Mariliis Öeren[1], PhD; Iain Jordan[1,2,3], MB, MSc, MRCPsych; Deborah Coughlin[1], MA; Sophie Turnbull[1,4], BSc, MSc, PhD

[1]Method X Studios Ltd, Sheffield, United Kingdom

[2]Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom

[3]Zinc VC, London, United Kingdom

[4]Academic Unit of Primary Health Care, University of Bristol, Bristol, United Kingdom

Corresponding Author:
Mariliis Öeren, PhD
Method X Studios Ltd
The Old Workshop, 1 Ecclesall Road South
Sheffield, S11 9PA
United Kingdom
Phone: 44 7722487501
Email: mt679@cantab.ac.uk

Abstract

Background: Although mental ill-health is more prevalent among people from lower socioeconomic groups, digital mental well-being innovations are often developed for people from higher socioeconomic groups, who already have resources to maintain good mental and physical health. To decrease health inequalities and ensure that available solutions are appealing and accessible to people with fewer resources, new approaches should be explored. We developed the app Wakey!, which focused on creating engaging mental health content that is accessible, particularly among lower socioeconomic groups in the United Kingdom.

Objective: The aim of this study is to assess engagement with the app, investigate initial effectiveness data for 6 well-being outcomes, and explore participants' subjective experiences of using Wakey!

Methods: The app Wakey! was publicly launched on January 20, 2020, and was free to download from Apple Store and Google Play. The app provided its users with entertaining and educational content related to mental well-being. Concurrently, a single-arm mixed methods feasibility trial was carried out from January to April 2020 among people who had downloaded the app and created an account. The primary outcome was engagement, which was collected passively from data logs. Secondary outcome measures were 6 well-being outcomes collected from self-report questionnaires. Individual interviews with 19 app users were carried out in April 2020.

Results: In total, 5413 people fit the inclusion criteria and were included in the final sample—65.62% (3520/5364) women, 61.07% (3286/5381) aged between 25 and 44 years, 61.61% (2902/4710) in employment, 8.92% (420/4710) belonging to the lower socioeconomic group, and 8.09% (438/5413) were engaged users. There was no evidence of a difference in engagement regarding sociodemographic and socioeconomic characteristics. There was evidence that users with a higher average daily sleep score, who joined the study more recently, who had higher baseline self-report of sleep quality, and who found episodes more entertaining were more likely to be engaged users. Among 230 users who provided follow-up data, there was evidence of improvements on four of the six well-being outcomes: life satisfaction ($P<.001$), feeling that life is worthwhile ($P=.01$), ease of getting up in the morning ($P<.001$), and self-efficacy ($P=.04$). The app and its content were well received by those who were interviewed, and several people perceived a positive change in their mental well-being.

Conclusions: This study shows that the app Wakey! could potentially be engaging across different socioeconomic groups, and there is an indication that it could positively impact the mental well-being of those engaged with the app. However, this study was a pragmatic trial with a limited sample, and the selection bias was present in the qualitative and quantitative study. Further work is needed to make any generalizable conclusions.

Trial Registration: ClinicalTrials.gov NCT04287296; https://clinicaltrials.gov/ct2/show/NCT04287296

Introduction

Background

Mental health conditions are a considerable burden for patients and health services and have been shown to have social patterning in severity and incidence [1,2]. People in the lowest socioeconomic groups have mental ill-health at higher rates than those in the highest socioeconomic groups [3,4]. Those in lower socioeconomic groups are more likely to be unemployed, working in jobs with low pay, and have insecure work, which have been found to be detrimental to mental health [3,5-7]. Mental health services are struggling to cope with demands on services, and unequal access to support is further exacerbating health inequalities [8,9]. Prevention and broader determinants of health have a larger effect on mental health than reactive, illness-based treatment [10,11]. In addition, people with higher education and higher socioeconomic background who have good access to resources (eg, time, income, and knowledge) are more likely to use commercial mental health solutions in the market to invest in their self-care [12-14].

Digital interventions have been proposed as a solution to address the high demand for mental health support in the context of the crisis in health care services [15]. During the past decade, there has been an explosion of available apps offering mental health and well-being support [16]. These apps target a variety of needs from habit formation to supporting recovery from mental ill-health [17]. Despite the large number of available apps, most lack evidence of effectiveness (ie, no available data) [18] and long-term engagement [19]. Results from a systematic search by Baumel et al [20] show how most apps see a drop in retention between days 1 and 30, 69% and 3%, respectively, depending slightly on the focus of the app (eg, happiness and meditation). Success in engagement has been shown to be a combination of several factors, such as higher rating in app stores, lower price, more positive reviews, good usability, variety in content and features, personalized experience, credibility, high security, social support, and the use of behavior change techniques (BCTs) [21-24]. McKay et al [17] found that 2 BCTs seemed to be more common to use in mental health apps, allowing or encouraging practice or rehearsal in addition to daily activities and providing instructions on how to perform the behavior.

The role of digital mental health interventions in addressing health inequalities is yet to be determined. On the one hand, they provide the potential to reduce health disparities, by providing personalized, low-cost, infinitely reusable resources that can increase access to health interventions [25,26]. On the other hand, they may increase inequity where there remain barriers to access and usability for disadvantaged groups [27,28]. To ensure that health inequalities are not further exacerbated by digital interventions, there is a need to develop and assess digital interventions that manage to maintain long-term engagement and that are appealing and accessible to people from lower socioeconomic groups who often use entertainment to regulate difficult emotions and for education [29-31].

We developed Wakey!, a mental well-being app that is generally appealing across social groups. However, we also aim to address inequalities in access to mental health and well-being support for lower socioeconomic groups in the United Kingdom, by providing content that is both entertaining and led by theory and evidence. The content was developed using information and techniques from cognitive and third-wave psychotherapies, positive psychology interventions, and mental health interventions [32-37]. The web-based psychoeducational strategies are effective at improving mental health literacy [33], reducing stigma [38], and improving the clinical course in depression [39]. Psychoeducational interventions in general are effective (with a small effect size) in managing stress [40].

Objectives

The purpose of this feasibility study is to explore engagement, assess initial impact, and explore users' subjective experiences with the app, to inform the next steps. Although we measured the app's impact on health outcomes, the study was not intended to be a definitive effectiveness trial.

Methods

Overview of Study Design

A 12-week mixed methods single-arm feasibility trial was conducted to explore engagement with the app Wakey!, initial effectiveness data, and subjective user experiences. Participants were recruited for the quantitative component exploring the engagement and effectiveness of Wakey! between January 17 and March 30, 2020. The qualitative study to explore people's experiences of using Wakey! was conducted between April 9 and 24, 2020.

Ethical Considerations

The trial was approved by the Faculty of Health Sciences Research Ethics Committee at the University of Bristol (reference 98382) and registered in ClinicalTrials.gov (NCT04287296).

Intervention—The App Wakey!

Wakey! is an app that delivers a 9-minute morning edutainment show, designed as an alternative to an alarm clock (ie, where users can set an alarm and wake up to a breakfast show on their phone). Edutainment refers to media where entertainment is combined with education [41]. The initial pilot of the show (10 episodes) was streamed via Facebook from April to May 2019. During this time, both quantitative and qualitative data (such as desired improvements) were collected and subsequently used as input in the development of the show and the app. Approximately 40.37% (44/109) of the sample had an annual household income below £30,000 (US $40,836.40), which suggests that the feedback received reflected the thoughts and

XSL•FO

RenderX

experiences of our target group. In addition, in February 2020, we carried out a think-aloud study among people who lived in a deprived area of London (Southwark and Lambeth). This provided us new insights that influenced usability tweaks in the app, which were implemented before the trial began.

This study focuses on the first 12 weeks after the launch on January 20, 2020, on Apple Store and Google Play. During this time, the show was presented by a comedy drag queen (Ginger Johnson) and a previous cast member from Love Island (Christopher Taylor), who were occasionally accompanied by special guests. Although including a drag queen as a host may turn off some lower socioeconomic group users who may hold more conservative social values, there is evidence that drag queens are seen as an important part in the working-class entertainment in the United Kingdom [42], and people from lower socioeconomic groups tend to watch Independent Television programs, such as Love Island, which is aimed at a less conservative audience [43].

Each week focused on a different mental health–related topic that was discussed by using colloquial language paired with humor in the morning shows (Multimedia Appendix 1). The content was developed (and written by one of the authors (IJ) who is a consultant psychiatrist and psychotherapist) as psychoeducational content designed to communicate the underlying principles from a variety of evidence-based treatments of, and preventative strategies for, common mental disorders. The content was composed of easy-to-understand principles from various cognitive therapies [44], third-wave psychotherapies such as compassion-focused and [45] mindfulness-based therapies [46], positive psychology interventions (such as well-being therapy developed by Fava [34]), and behavioral and lifestyle mental health interventions such as sleep behavior, exercise, and behavioral activation [47]. The content was adapted for the scripts for the show in collaboration with one of the hosts of the show who is a professional writer. Additional content and features were provided on the app to support the theory- and evidence-based techniques, to increase engagement with the app, and to respond to current events (such as the COVID-19 crisis; Multimedia Appendix 1).

Recruitment

The app was advertised in the United Kingdom through partner companies of the *Lost In TV Audience Services*, who have a database of 450,000 people (approximately 10% were targeted). In addition, the app's content was advertised through social media platforms (eg, Facebook and Instagram) and more traditional media outlets (eg, *Marie Claire*, *Sunday Mirror*, *The Sun*, and *Metro*).

As this was a feasibility study and focused primarily on uptake and engagement with an interest in the demographics of the users, a sample size calculation was not performed.

Quantitative Study

Procedures

On downloading the app, users were taken through a registration process. Users were presented with the Terms and Conditions

and the Privacy Policy, which included consent to enter the study and the data to be used for research purposes (Multimedia Appendix 1). Thereafter, users were asked to provide sociodemographic and socioeconomic information: name, email address, age range, gender, and occupation (Multimedia Appendix 1). Users were asked to set their in-app alarm clock time. Baseline data were then collected on the 4 UK Office for National Statistics (ONS) well-being questions [48], one question inquiring about ease of waking up in the morning and one question inquiring about self-efficacy (Multimedia Appendix 1). Following sign-up, users were taken to the home page (Multimedia Appendix 1), where they were presented with the welcome video.

Data Collection

All quantitative data were collected from users via the app. The data on overall engagement were collected passively from data logs on a daily basis (ie, whether users watched that day's episode or not). Other measures were collected by asking users to provide information at baseline (ie, onboarding) and thereafter on either a daily or weekly basis. Sociodemographic and socioeconomic information was collected only at baseline. A total of 6 well-being outcomes were collected at baseline and then weekly until the end of the trial. Engagement outcomes included the number of people who downloaded the app and created an account, number of shows watched over the 12-week period, average time watched, and entertainment value of the episode (self-reported on a scale of 0 [*not at all*] to 10 [*completely*]).

Data Processing

Users were excluded from all analyses if they had not finished creating the account or were aged <18 years. Users whose baseline and follow-up scores were left on the default setting were excluded from the impact assessment. This was 0 for the ONS and sleep questions and *strongly disagree* for the self-efficacy question. In addition, when participants answered *Prefer not to say* to gender (n=49), age (n=32), or occupation (n=703) questions, their answer for a specific variable was treated as missing. Owing to a technical error, 63 users were not able to answer the self-efficacy questions when creating the user, thus missing the baseline assessment and excluded from the analysis about self-efficacy.

Users were segmented into five levels of engagement: never active—had not seen any of the episodes and the welcome video; inactive—had seen only 1 episode or the welcome video or both; became inactive after their first week—saw at least two episodes on their onboarding week and then stopped watching; irregular—had seen at least two episodes on separate weeks but <20% of all the available episodes for them; and engaged users—had seen ≥20% episodes of those available to them. The 20% threshold was a rough equivalent of weekly use—if the user would watch 1 episode per week, then it would equal to 20% of weekly episodes.

The coding of occupational groups is based on two classifications: the Standard Occupational Classification 2010 volume 1 [49] and 2016 ONS National Statistics Socio-economic Classification (NS-SEC) [50]. The 8 NS-SEC

categories provided in Table S1 in Multimedia Appendix 1 were used to provide a more detailed overview of socioeconomic groups within the sample of people who used Wakey! during the 12-week trial in the descriptive analysis and to explore the predictors of engagement with Wakey! and improvements on the well-being outcomes.

Data Analysis

To see if there were any differences in the probability distributions between the active and never active groups and between people who provided follow-up data and those who did not by sociodemographic and socioeconomic characteristics and content use characteristics (ie, number of watched episodes and entertainment value), chi-square tests and independent-group 2-tailed t tests (or Wilcoxon rank-sum tests) were conducted. Wilcoxon signed-rank tests were used to investigate improvements in the 6 ordinal well-being outcomes from baseline to follow-up. To explore predictors of outcomes, binary variables were created, and multiple logistic regression was undertaken, as the parameters were not met for linear regression. For engagement, we explored which variables predicted whether the participant was an engaged user (watched ≥20% of episodes available) or a not engaged user (watched <20% of episodes available). For retention, both the daily (ie, days 1, 7, 14, and 30) and weekly (weeks 1-5) retention were assessed by calculating the proportion of users who created an account and then were active at a specific time. For well-being outcomes, we explored predictors of improvements (≥1 point change) versus no improvement (no change or deterioration). Univariable analysis was conducted using logistic regression to explore predictors of whether the participant was an active user or improved on well-being outcomes. All predictor variables are presented in Table S1 (Multimedia Appendix 2). Repeated-measures multivariable analysis was performed using logistic regression. The initial model inclusion criterion was $P<.05$, with putative predictors entered using backward stepwise selection and retained where $P<.05$. To explore the influence of exposure to Wakey! and to account for different entry times into the study for different participants, we explored associations between improvements in well-being outcomes and user segmentation, the week when they joined the study, if they were an active user, the time between baseline and the last follow-up, and the last week they provided follow-up data.

Qualitative Study

Procedures

The aim was to interview people from lower socioeconomic groups [49,51] and from a diverse range with respect to gender, age range, and user engagement with Wakey! We used purposive sampling [52] and divided people into different groups based on their engagement (to ensure that people from all 4 groups would be represented in the study) and then sent invitations by email (at first) and push notifications to participate

in the study (Multimedia Appendix 1). People who were interested in participating in the study received an email with the participant information sheet and a link to the web-based consent form. The web-based consent forms were hosted on the University of Bristol BOS system, and the data were kept in accordance with the Data Protection Act 2018 [53].

Data Collection, Processing, and Analysis

Qualitative data were collected via semistructured audio-recorded individual interviews conducted by MÖ. All interviews occurred either on the phone or on a video-call platform. Participants were given a £20 (US $27.22) high-street voucher if they were an active user and £5 (US $6.81) if they were an inactive user as a thank-you for their time. The interview topic guide is outlined in Multimedia Appendix 1. The audio recordings were made on encrypted audio-recorders and transferred to Method X Studios secure servers, where they were kept in accordance with the Data Protection Act 2018 [53]. The anonymized transcriptions were kept separately from the identifiable information on the consent forms, so they could not be linked. Transcribed recordings were anonymized (all names or other identifying material removed), and the collected data were analyzed in themes, which were based on the interview topic guide (using a deductive approach). Gender, age range, and user group have been added to all quotes presented in the *Results* section.

Results

Quantitative Study Sample

Between January 17 and March 30, 2020, a total of 5928 people downloaded Wakey! (unique downloads, excluding the Wakey! Team). Of these 5928 people, 515 (8.69%) did not meet the inclusion criteria and were excluded from the study (276 were aged <18 years and 239 had not verified their email address and thus did not finish the registration), leaving a final sample of 5413 users, who were divided into two groups—never active and active. The characteristics of active and never active users are presented in Table 1. Two-thirds of the active users were women (3520/5364, 65.62%) and aged between 25 and 44 years (3286/5381, 61.07%). Approximately 61.61% (2902/4710) of the users were in employment. Approximately 8.92% (420/4710) of the users had an occupation that indicated belonging to a lower socioeconomic group (as defined by the NS-SEC [50]), such as semiroutine and routine occupations (Table S2 in Multimedia Appendix 1). A higher proportion of the never active group was women, in the youngest age group, and not working (unemployed, caregivers, retired, students, looking after family or home, or sickness or disability). In terms of socioeconomic groups, those who were never active were also more likely to have an occupation that indicated them belonging to lower socioeconomic groups.

Table 1. Sociodemographic and socioeconomic characteristics among active and never active users.

	Active, n (%)	Never active, n (%)	*P* value
Gender (active N=3958, never active N=1406)			<.001
Female	2489 (62.89)	1031 (73.33)	
Male	1404 (35.47)	358 (25.46)	
Nonbinary or other	65 (1.64)	17 (1.21)	
Age (years; active N=3972, never active N=1409)			<.001
18-24	643 (16.19)	300 (21.29)	
25-34	1268 (31.92)	459 (32.58)	
35-44	1195 (30.09)	364 (25.83)	
45-54	689 (17.35)	236 (16.75)	
55-64	153 (3.85)	42 (2.98)	
≥65	24 (0.60)	8 (0.57)	
Occupation (active N=3455, never active N=1255)			<.001
Employed	2191 (63.42)	711 (56.65)	
Unemployed	137 (3.97)	56 (4.46)	
Caregivers	241 (6.98)	94 (7.49)	
Retired	53 (1.53)	10 (0.80)	
Students	363 (10.51)	184 (14.66)	
Looking after family or home	301 (8.71)	132 (10.52)	
Sickness or disability	169 (4.89)	68 (5.42)	
Socioeconomic groups (active N=3455, never active N=1255)			<.001
High	1462 (42.32)	438 (34.90)	
Middle	414 (11.98)	136 (10.84)	
Low	290 (8.39)	130 (10.36)	
Not classified elsewhere	1289 (37.31)	551 (43.90)	

User Retention and Engagement

Among all 5413 users (this includes both active and nonactive users), 1593 (29.43%) were active on day 1, a total of 273 (5.04%) on day 7, 169 (3.12%) on day 14, 126 (2.33%) on day 21, and 108 (2%) on day 30.

As Wakey! was meant to be used during weekdays and people were onboarded during weekends, it was decided to assess the weekly retention as well. Of the 5413 people, 3135 (57.93%) were active on the week they were onboarded, 1454 (26.87%) 1 week later, 635 (11.73%) 2 weeks later, 448 (8.28%) 3 weeks later, and 341 (6.30%) were active 4 weeks later.

The welcome video (length 1 min 26 seconds) was seen by 24.46% (1324/5413) of the users, with a mean watch time of 1 min 6 seconds. The mean number of users watching live streamed episodes was 127 (SD 83.67), and that for archives episodes was 102 (SD 92.22). When combining the unique views of livestream and archived episodes, the mean number of viewers was 219 (SD 131.65). The mean entertainment value of all 60 episodes was 6.80 out of 10 (SD 1.14). Compared with the first week (mean 5.64, SD 0.68), the average score increased by 2.34 points by the last week (mean 7.98, SD 0.43).

When dividing the 5413 users based on the level of engagement, 1420 (26.23%) were never active, 2024 (37.39%) were inactive users, 406 (7.50%) became inactive after their first week, 1125 (20.78%) were irregular users, and 438 (8.09%) were engaged users. The engaged users were divided into those who saw 20%-39% of the episodes available to them (269/5413, 4.97%), 40%-59% (83/5413, 1.53%), and 60% or more of the episodes available to them (86/5413, 1.59%).

In the univariable analysis (excluding the never active users), being engaged users was predicted by a higher average entertainment score (odds ratio [OR] 1.16, 95% CI 1.12-1.20; *P*<.001), a higher average daily sleep score (OR 1.15, 95% CI 1.11-1.20; *P*<.001), joining the study more recently (OR 0.92, 95% CI 0.88-.96; *P*<.001), and a higher baseline report of sleep quality (OR 1.06, 95% CI 1.03-1.10; *P*=.001) (Table S2 in Multimedia Appendix 2). There was no evidence of a difference in engaged users (watched ≥20% of available episodes) versus not engaged users (watched <20% of available episodes) in terms of social characteristics (Table S3 in Multimedia Appendix 2). The results from the multivariate model are presented in Table S4 (Multimedia Appendix 2).

Effectiveness on Well-being Outcomes

When comparing users (N=3993) who had seen at least one episode (or the welcome video) and provided follow-up data with those who had seen at least one episode (or the welcome video) and had not provided any follow-up data, there was a higher proportion of users aged ≥45 years and users who were from middle socioeconomic groups and who provided follow-up data on ONS, sleep, and self-efficacy measures (Table S5 and Table S6 in Multimedia Appendix 2). In addition, users who provided follow-up data had rated episodes with a higher score

(mean=7.5 vs 5.2; $t_{1348}=-8.6134$, $P<.001$) and had watched more live (11 vs 1; $t_{3885}=-36.1515$, $P<.001$) and archived episodes (6 vs 1; $t_{3885}=-26.9019$, $P<.001$).

Users who provided follow-up data demonstrated improvements on 4 of the 6 health outcomes (Table 2). There was strong evidence ($P<.001$) that users experienced improvements on life satisfaction and ease of getting out of bed in the morning (median difference 2 points). There were also improvements in feeling that life is worthwhile (median difference 1, $P=.01$) and in self-efficacy (median difference 0, $P=.04$).

Table 2. Evidence of improvements in outcomes from baseline to final follow-up (N=5413).

	Users whose score improved		Users with baseline data, n (%)	Median at baseline (range, IQR)[a]	Users with follow-up data, n (%)	Median at follow-up (range, IQR)[a]	Difference from baseline to follow-up	P value
	N	n (%)						
Life satisfaction	230	112 (48.7)	3638 (67.21)	6 (0-10, 4-7)	230 (4.25)	7 (0-10, 5-8)	+1	<.001
Worthwhile	230	111 (48.3)	3638 (67.21)	6 (0-10, 4-8)	230 (4.25)	7 (0-10, 5-8)	+1	.01
Happy yesterday	230	97 (42.2)	3638 (67.21)	5 (0-10, 4-8)	230 (4.25)	6 (0-10, 4-8)	+1	.43
Anxious yesterday	230	97 (42.2)	3638 (67.21)	5 (0-10, 2-7)	230 (4.25)	4 (0-10, 2-7)	−1	.73
Easy to get up	230	133 (57.8)	3638 (67.21)	4 (0-10, 2-7)	230 (4.25)	6 (0-10, 4-8)	+2	<.001
Self-efficacy	158	53 (33.5)	3577 (66.08)	4 (1-5, 3-4)	158 (2.92)	4 (1-5, 3-4)	0	.04

[a]Baseline and final follow-up data were collected at different time points for different users, as people were able to sign up at any point during the 12-week trial, and follow-up data were collected each week. Follow-up data were reported as the final outcome data reported by the users.

There was a single variable that predicted improvements in life satisfaction ($P<.05$) in the univariable analysis; therefore, no multivariable analysis was conducted (Table S7 in Multimedia Appendix 2). Lower self-efficacy at baseline was associated with improvement in life satisfaction by the end of the trial (OR 0.63, 95% CI 0.50-0.80; $P<.001$).

There were 2 variables associated with improvements in feeling life is worthwhile by the end of the trial. Those participants who improved on perceiving the life worthwhile were more likely to have a lower baseline sleep quality (OR 0.89, 95% CI 0.81-0.97; $P=.01$) and have had lower self-efficacy at baseline (OR 0.78, 95% CI 0.63-0.98; $P=.03$). The results from the multivariate model are presented in Table S8 (Multimedia Appendix 2).

Four variables were associated with improvements in sleep quality in the univariable analysis: higher number of archived episodes watched (OR 1.04, 95% CI 1.00-1.08; $P=.04$) and lower baseline scores for three of the ONS questions—satisfaction with life (OR 0.83, 95% CI 0.74-0.93; $P=.002$), life being worthwhile (OR 0.77, 95% CI 0.69-0.87; $P<.001$), and happiness (OR 0.85, 95% CI 0.76-0.94; $P=.003$).

A single variable was associated with self-efficacy in the univariable analysis; those with a lower baseline score of perceiving life being worthwhile (OR 0.86, 95% CI 0.76-0.99; $P=.03$) were more likely to show improvements in self-efficacy.

Qualitative Study Sample

In total, 1524 people received an invitation to participate in the qualitative study, and of these 1524 people, 76 (4.99%) were

interested in participating. Ultimately, 1.25% (19/1524) of the users consented, all of whom were interviewed. Of the 19 participants, 9 (47%) were women, 12 (63%) were aged between 25 and 44 years, and 15 (79%) were not in full-time employment (eg, student, caregiver, and unemployed). On the basis of the segmentation of the level engagement, of the 19 participants, 9 (47%) were active users, 5 (26%) were irregular users, 2 (11%) had become inactive, and 3 (16%) were inactive users.

Qualitative Results

Hearing About and Using the App

Participants mainly heard about Wakey! through social media, such as Facebook and Instagram. When asked about what made the interviewees interested in the app, different reasons were mentioned, such as having a drag queen as one of the hosts, having a different approach to waking up, and addressing mental health issues (almost all interviewees reported having mental ill-health currently or in the past).

When asked about how they used the app, most used Wakey! as part of a morning routine after waking up or sometimes later during the day, rather than as an alarm clock. Whereas some people had seen change in their use of Wakey! during lockdown, others had stuck to their routine or way of life.

> *There's not much difference for me, to be honest, because I'm isolated anyway...There are a lot of people with physical disabilities and mental disabilities and illnesses that live like this already.* [Female, aged 35-44 years, active]

XSL•FO

RenderX

The interviewees who were active users said that they watched Wakey! on most mornings or every day (including watching old episodes during the weekend or watching archived episodes if they missed the livestream ones). When asked why they might have missed some of the episodes, the main reasons were the change in their morning routine due to lockdown (eg, change in work times and not having to wake up at a particular time), forgetting to watch the show, and occasionally oversleeping or staying up late. The use of the app varied slightly among the irregular users. Of them, 4 were watching the show at least three times a week, including occasional watch of the archived episodes. Getting late to bed or forgetting about the show were two reasons behind not watching the show more frequently. One of the irregular users felt that sometimes he feels it was too overwhelming to watch the show.

> *To be honest with you, it's how I feel on the day...It is basically about trying to be in the right state of mind...trying to be in the mood to watch Wakey Wakey is one thing because sometimes if you're not feeling too great within yourself, it's like the last thing you want is sometimes to be looking at some people, where everyone's laughing, because it's hard.* [Male, aged 25-34 years, irregular]

Watching Wakey! was usually less regular among people who became inactive after their first week or who had been inactive before the qualitative study. Some of them mentioned that they tend to watch Wakey! a couple of times a week or as often as possible. The reasons behind not watching the show more often were related to the change in their routine during lockdown (as among active users) and work routine.

The Presenters and Content

Interviewees' general impression of the app was positive. They were very satisfied with the presenters, and it was mentioned several times that they seem very friendly and have a great chemistry, which makes the show very enjoyable. Several people pointed out that the content of Wakey! was very well developed, as it covered a variety of useful topics on mental health and provided practical tips that people can relate to and incorporate into their lives, and it was done in a fun and entertaining way that helped keep people engaged.

> *You've wrapped it up in an entertainment bubble, but it's very much about mental health. I think it's brilliant the way it's done.* [Female, aged 35-44 years, active]

> *I think it's really good that they're focusing on a lot to do with mental health because that is a big thing that gets pushed under the carpet. There's so many of us that look okay on the outside but might be suffering.* [Female, aged 35-44 years, irregular]

Use of Features

Several interviewees had used the chat option during livestream episodes, Q and A sessions, or *Brain–aerobics* quizzes. They liked that using the chat created the feeling of a community through interacting with other users. In addition, it was appreciated that people received responses to their questions and comments.

> *I do appreciate the fact that somebody from Wakey! has actually replied to me and had a bit of a conversation with me. I do like the interactive aspect and getting involved with people."* [Female, aged 25-34, active]

By the time the interviews occurred, 8-9 articles were made available on the app. Whereas some people had read the articles and found them useful, most had not yet engaged with them.

There were two types of live events: Q and A and *Brain-aerobics* quiz. Not everyone was aware of the quiz, but those who were viewed it positively, describing it as an opportunity to interact with Ginger Johnson and other users. Only a small number of interviewees had watched Q and A live events (as people had other responsibilities at the time it was broadcast). Of those who had, some found them useful, whereas others did not feel they were relevant to them or found them boring.

Among users who had visited the progress page, the regularity of use varied; however, in general, people found it useful as a tool to track their progress and see improvements.

Data Collection

Daily questions at the end of episodes were perceived, among some participants, as a possibility for self-reflection. Whereas some of them would have liked more daily questions, others found them too frequent. Whereas the majority remembered the daily questions, several struggled to remember if they had answered the weekly ones. Those who had answered the questions commented that the questions can be useful when you are interested in your mental well-being and show how things have changed compared with previous weeks.

> *When you actually put that number in on the score, you think to yourself, "I actually feel a lot better than what I thought," or, "I actually feel a lot worse than what I thought." I think it's a good idea definitely because it helps well, yourself to be able to see how you are feeling.* [Male, aged 25-34 years, irregular]

> *When I joined, I was quite shocked by doing the survey at the end, when you have to give yourself points as to how you feel and things. It shocked me how bad I was feeling. I hadn't realized that that's how I was feeling.* [Female, aged 55-64 years, active]

Perceived Change in Outcomes

When asked about changes in mood, sleep, or other similar aspects that might relate to using Wakey!, almost all interviewees had some positive examples of change that related to improvement in their mood, paying more attention to having a routine (which was perceived as likely to have an effect on sleep quality), finding it easier to getting up from the bed in the mornings, and reduced stress levels.

> *Using the app made me feel better in the morning, got me up, got me alert, got me awake, relatively easy. Whereas I previously tended to snooze my alarm clock frequently.* [Male, aged 45-54 years, irregular]

> *It just gets me out of bed at a proper time. I've noticed that in personally myself, I've been able to smile a*

bit more rather than being a bit down in the dumps. Even if I am in a bit of a mood when I wake up, by the time I've watched an episode of Wakey!, I've giggled me up for a good 10 minutes. [Male, aged 35-44 years, active]

Usability Issues and Future Improvements

Although people had many positives to say about the content of Wakey!, there were things that participants felt could be improved, such as some users feeling that the show felt a bit rushed and too preplanned and some of its content repeating itself. Technical issues mentioned included the following: alarm not going off, videos freezing, issues with chat's functionality, livestream episodes not starting from the beginning, and not being able to save the videos or watch them offline.

When interviewees were asked about three things they would keep if everything else about Wakey! would be changed, the three most popular things were the presenters and the Old News and *Bed-Aerobics* segments from the morning shows (described in Multimedia Appendix 1).

That's a tough one. Things to stay, obviously, first of all, is the presenters. Secondly, I'd say the content...is fantastic. [Male, aged 25-34 years, active]

Several participants found it difficult to answer but instead focused on things that could be improved. Suggestions included addressing different parts of the app, such as the content, survey, chat, live events and challenges, length and frequency of the episodes, and technical issues. One of the interviewees emphasized that the app's objective is not entirely clear and could have an effect on people not using it.

I think one of the main things is it's not entirely clear what the app's meant to do. It's a wake-up alarm clock, and the mental health and well-being side of it's not that clear until you've started using it. [Male, aged 45-54 years, irregular]

Discussion

Principal Findings

This study has three main objectives: (1) exploring engagement, (2) assessing the app's potential to improve mental well-being, and (3) exploring users' subjective experiences with the app. The findings provide insights into the potential an edutainment app can have, from both the engagement and effectiveness perspectives. Fleming et al [19] have shown that the use of digital interventions can be more modest when launched in a real-world setting (compared with a trial setting). As Wakey! was made publicly available to everyone, the findings are more likely to reflect the *real-world* data and to be ecologically valid. Although 8.09% (438/5413) of the users were considered as engaged, it is difficult to compare the level of activity at Wakey! with other apps, owing to differences in definitions and the content [19]. For example, moderate use can be viewed as using an app 1 week after installation, but it can also mean that a proportion of users (7%-16%) have completed at least two modules on the app [19]. In contrast, there are universal indicators such as user retention that enable the initial comparison. Baumel et al [20] showed that, on average, around

70% of mental health apps' users are active on day 1. That number drops significantly and reaches 10% by day 7 and 4% by day 30. When comparing the user retention numbers of Wakey! with the average, the former are around twice as low at each time point. At the same time, Wakey! is a weekday app, with no new content added during the weekend, and as the expectation for an active user was to use the app at least once a week, it was decided to see the weekly retention as well. Although the week 1 retention of Wakey! is still lower than the average day 1 retention, user activity does not drop that rapidly in the following weeks. Although days 1, 7, 14, and 30 retention rates tend to be more common to use [20], the authors would recommend considering using weekly retention as an accompanying metric, especially if the app is not meant for daily use.

Although Wakey! aims to decrease inequalities by increasing access to mental health information and products to people from lower socioeconomic groups, only a small proportion of users were from these groups. In addition, people who had never used the app were more likely to be from lower socioeconomic groups. Taken together, these findings suggest that the version of Wakey! explored in this trial was less accessible or appealing to people from lower socioeconomic groups. Low uptake can be influenced by a variety of factors, such as low digital literacy, lack of awareness of the app, low availability and accessibility, lack of recommendations from other people to use the app, and lack of support to navigate new technology [54,55]. The difference in use has been explored in several systematic reviews, but the evidence regarding socioeconomic characteristics is inconclusive. For example, Turnbull [56] found that people with higher income were more likely to use digital solutions that addressed chronic health conditions, but there were no differences in use regarding education and employment. However, it was highlighted that caution should be taken with conclusions drawn from these findings because the number of studies included in the analysis was small and there was a high risk of bias. Beatty and Binnion [57] investigated the adherence to web-based psychological interventions and found that in 28% of studies reporting education, higher adherence was predicted by higher education. Similar to Turnbull [56], the authors did not find any association between employment and adherence in all studies reporting employment. Perski et al [58] found that higher levels of education and employment were associated with greater engagement with digital behavior change solutions.

On the basis of the responses from a small number of users who provided qualitative data, participants' general experience with the app was positive. This was in agreement with the high average entertainment scores reported in the quantitative data. The latter, in turn, was one of the indicators that predicted higher engagement with the app. The results of this study also indicate that in the small sample (n=158-230) of users who were engaged with the app and provided data for health outcomes, improvement was seen across different mental well-being outcomes, such as life satisfaction, life being worthwhile, ease of waking up and self-efficacy, and users' perception of positive change in their mental well-being. A recent systematic review highlighted that only 2% of publicly available psychosocial

wellness and stress management mobile apps have published peer-reviewed evidence of feasibility or efficacy [18].

Mental health apps vary by different factors, such as the mental health conditions they target [59,60] and the BCTs used [17]. For example, providing instructions on how to perform a behavior and allowing or encouraging practice or rehearsal in addition to daily activities are the two most popular BCTs used in apps that aim to improve mental well-being [17]. The content of Wakey! has influences from different types of therapies and psychology interventions and covers a variety of BCTs shown to have an impact on mental well-being [17,61]. Although the concept that Wakey! uses—mixing entertainment with education—has been used for decades in various forms (eg, *Sesame Street* in the Unites States and *The Archers* in the United Kingdom) [62], its uniqueness largely stems from its format—it is about addressing mental well-being at the start of the day through a breakfast show. However, the app should be more appealing to the target population who already tends to be less likely to use digital health solutions [63-65].

There are several ways to make the app more appealing to people from all socioeconomic groups. Our qualitative findings in this study suggested that not everyone was aware of the app's objective when they downloaded it (ie, a focus on improving mental well-being) and providing a clear description of the app in social media and in the App Store and Google Play is "a key channel to inform consumer choice" [66]. For some people, this led to dissatisfaction and disengagement because they were seeking entertainment, not entertainment and support. We will seek to clarify the description of the app in the App Store and Google Play, to improve clarity of the purpose of the app, and to reduce misunderstanding about its purpose. To make the app more appealing and user-friendly to lower socioeconomic groups specifically (as we did not quite reach the groups expected), further work (eg, cocreation workshops and feedback sessions) is being undertaken with the target group. These methods are evidenced to improve the accessibility of behavior change interventions and to improve engagement in these groups [67,68].

Huang and Bashir [21] investigated how information cues across anxiety apps influence the selection and adoption of the app. They found that users were more likely to select apps that were cheaper and had better ratings and reviews. Schueller et al [69] asked people to rate the importance of health apps' features and found that content, ease of use, and cost were the most important factors regarding uptake and continued use. Alqahtani and Orji [22] found that poor usability, unvaried content, and lack of personalization were the most common reasons why users stop using mental health apps. Although the general experience of the app described by the sample of qualitative study participants (n=19) was positive (eg, easy to navigate, useful, and entertaining content), it was also pointed out that technical issues should be addressed, and the content could be more varied. Although there is a lack of information regarding the main reasons why people stopped using the app, several interviewees mentioned the issue of the alarm not working properly, and similar feedback was also received via emails and morning chat. Thus, our best guess is that functional issues with the alarm could be one of the main reasons why people stopped using the

app (ie, becoming inactive). The app was advertised as an alternative to a traditional alarm clock, and the feedback received implied that it was not fully functional for everyone—the phone had to be unmuted and connected to the internet.

Uptake and ongoing use of apps are influenced by a variety of factors, which are often related to users' needs and resources. To reduce barriers to access for those with lower incomes, Wakey! is designed to be free to use and the entertaining content becomes available as soon as the user creates an account.

Limitations

First, the study lacked the comparator group. It is therefore not possible to know if improvements in the 4 well-being outcomes were related to the use of the app.

Second, the sample size of people who provided follow-up data is small, and this limits the generalizability of the findings. Although there was an indication that these users improved on 4 health outcomes, these results should be interpreted with caution as the data are likely to be not missing at random [70]. Users who provided health data had rated episodes with a higher score and had watched more episodes; a higher proportion were aged ≥45 years and were more likely to be from a middle socioeconomic group. Therefore, this group of individuals was not representative of the whole sample but can be taken of an indication that those engaged with the app can benefit from its use. Using incentives (eg, quizzes and prizes) would be an option to motivate people who do not normally engage in research to provide data.

Third, the app was targeting people from lower socioeconomic groups, but the group was underrepresented in the final sample (420/4710, 8.92%).

Fourth, more than one-third of users (1840/4710, 39.07%) who provided data on their occupation belonged to a group *Not classified elsewhere*. This group included people who were not in employment—unemployed, caregivers, retired, students, looking after family or home, or sickness or disability. It is possible that a proportion of these people who had engaged with Wakey! would have been classified as people from lower socioeconomic groups, as those people who are not in employment may have lower levels of access to resources. However, because we did not have any information about people's income or prior occupations, it was not possible to infer their socioeconomic group in this study. In a future study, prior occupations and other indicators of access to resources would be sought to support the identification of the socioeconomic group.

Fifth, the frequency of follow-up reports varied among the participants, from 35.2% (81/230) of the people whose last follow-up was after 1 week of using Wakey! and 5.2% (12/230) from the 12th follow-up week. To account for this, we explored the influence of study entry dates and the dates of follow-up data in the analysis of predictors of improvements in outcomes. As only a small number of variables were captured to reduce the data burden on the app's users, it is likely that several important covariates that affected participants' mental

well-being were not captured, including social characteristics (such as ethnicity) and mental health indicators.

Sixth, 3 of the 6 well-being questions asked were based on recent past (how happy and anxious you were yesterday and how easy it was to get up this morning) and might not reflect users' overall state.

Seventh, there was a lower proportion of users who were less engaged with the app (inactive or who had become inactive user groups) in the qualitative study compared with those who were more engaged (ie, active and irregular user groups). Although we used a diverse range of recruitment methods (eg, sent emails and push notifications and used incentives) to include an equal number of participants from each group, we found it very challenging to get inactive users to engage with the qualitative study. Ultimately, we interviewed everyone who expressed interest by the end of the study, thus potentially introducing self-selection bias [71]. We appreciate that this limits the conclusions that can be drawn from this study as we did not reach data saturation and interviewed those who were more motivated to share their experience. Interviewing more inactive people might have provided us some additional insights that potentially may have diverged from our sample.

Eighth, the occurrence of the COVID-19 lockdown may have had an impact on the study outcomes and user engagement with the app.

Conclusions

Digital mental well-being solutions are often aimed at people from higher socioeconomic groups, and the majority struggle with high drop-off rates. This study shows that an app could not only be potentially engaging across socioeconomic groups if its content is grounded in theory and evidence but also be engaging and entertaining. There is also an indication that this type of app can have a positive impact on mental well-being among more engaged users.

However, this study was a pragmatic trial, which was based on a limited sample without a control group, and these results apply to the particular group of people in the study. Thus, it is not possible to generalize the results to a wider population.

As we saw that the uptake of the app was significantly lower in lower socioeconomic groups, further work that involves cocreation with the target group is to be undertaken.

Authors' Contributions

ST, IJ, and DC were involved in the study design. ST and MÖ carried out the study, analyzed the data, and drafted the manuscript. All authors read and contributed to the finalization of the manuscript. All authors approved the final manuscript.

Multimedia Appendix 1
Supporting information: the app, socioeconomic groups, qualitative study.
[DOCX File , 2493 KB - formative_v6i3e25715_app1.docx]

Multimedia Appendix 2
Supporting tables (sociodemographic and socioeconomic characteristics, variables included in analysis, multivariate logistic regression models, and univariable associations).
[DOCX File , 47 KB - formative_v6i3e25715_app2.docx]

References

1. Jokela M, Batty GD, Vahtera J, Elovainio M, Kivimäki M. Socioeconomic inequalities in common mental disorders and psychotherapy treatment in the UK between 1991 and 2009. Br J Psychiatry 2013;202:115-120. [doi: 10.1192/bjp.bp.111.098863] [Medline: 22500010]
2. The global burden of disease: 2004 update. World Health Organization. 2008. URL: https://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf?ua=1 [accessed 2020-07-14]
3. Marmot M, Bell R. Fair society, healthy lives. Public Health 2012;126 Suppl 1:S4-10. [doi: 10.1016/j.puhe.2012.05.014] [Medline: 22784581]

4. Health matters: reducing health inequalities in mental illness. Public Health England. 2018. URL: https://tinyurl.com/nhc3bsf7 [accessed 2020-06-22]

5. Campion J, Bhugra D, Bailey S, Marmot M. Inequality and mental disorders: opportunities for action. Lancet 2013;382(9888):183-184. [doi: 10.1016/S0140-6736(13)61411-7] [Medline: 23870520]

6. Stansfeld S, Smuk M, Onwumere J, Clark C, Pike C, McManus S, et al. Stressors and common mental disorder in informal carers--an analysis of the English Adult Psychiatric Morbidity Survey 2007. Soc Sci Med 2014;120:190-198. [doi: 10.1016/j.socscimed.2014.09.025] [Medline: 25259657]

7. Stewart-Brown S, Samaraweera PC, Taggart F, Kandala NB, Stranges S. Socioeconomic gradients and mental health: implications for public health. Br J Psychiatry 2015;206(6):461-465. [doi: 10.1192/bjp.bp.114.147280] [Medline: 25792696]

8. Gilson L, Doherty J, Loewenson R, Francis V. Challenging inequity through health systems: final report knowledge network on health systems. World Health Organization. 2007. URL: https://www.who.int/social_determinants/resources/csdh_media/hskn_final_2007_en.pdf [accessed 2020-07-14]

9. Breaking point: the crisis in mental health funding. Trades Union Congress. 2018. URL: https://www.tuc.org.uk/sites/default/files/Mentalhealthfundingreport2_0.pdf [accessed 2020-08-27]

10. Arango C, Díaz-Caneja CM, McGorry PD, Rapoport J, Sommer IE, Vorstman JA, et al. Preventive strategies for mental health. Lancet Psychiatry 2018;5(7):591-604. [doi: 10.1016/S2215-0366(18)30057-9] [Medline: 29773478]

11. McGinnis JM, Williams-Russo P, Knickman JR. The case for more active policy attention to health promotion. Health Aff (Millwood) 2002;21(2):78-93. [doi: 10.1377/hlthaff.21.2.78] [Medline: 11900188]

12. Azzopardi-Muscat N, Sørensen K. Towards an equitable digital public health era: promoting equity through a health literacy perspective. Eur J Public Health 2019;29(Supplement_3):13-17 [FREE Full text] [doi: 10.1093/eurpub/ckz166] [Medline: 31738443]

13. Bol N, Helberger N, Weert JC. Differences in mobile health app use: a source of new digital inequalities? Inf Soc 2018;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]

14. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]

15. Castle-Clarke S. What will new technology mean for the NHS and its patients?: four big technological trends. The Kings Fund. 2018. URL: https://www.kingsfund.org.uk/sites/default/files/2018-06/NHS_at_70_what_will_new_technology_mean_for_the_NHS_0.pdf [accessed 2020-02-16]

16. Aitkin M, Clancy B, Nass D. The growing value of digital health in the United Kingdom: evidence and impact on human health and the healthcare system. IQVIA Institute for Human Data Science. 2017. URL: https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/the-growing-value-of-digital-health.pdf [accessed 2020-05-24]

17. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using health and well-being apps for behavior change: a systematic search and rating of apps. JMIR Mhealth Uhealth 2019;7(7):e11926 [FREE Full text] [doi: 10.2196/11926] [Medline: 31274112]

18. Lau N, O'Daffer A, Colt S, Yi-Frazier JP, Palermo TM, McCauley E, et al. Android and iPhone mobile apps for psychosocial wellness and stress management: systematic search in app stores and literature review. JMIR Mhealth Uhealth 2020;8(5):e17798 [FREE Full text] [doi: 10.2196/17798] [Medline: 32357125]

19. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]

20. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]

21. Huang HY, Bashir M. Users' adoption of mental health apps: examining the impact of information cues. JMIR Mhealth Uhealth 2017;5(6):e83 [FREE Full text] [doi: 10.2196/mhealth.6827] [Medline: 28659256]

22. Alqahtani F, Orji R. Insights from user reviews to improve mental health apps. Health Informatics J 2020;26(3):2042-2066 [FREE Full text] [doi: 10.1177/1460458219896492] [Medline: 31920160]

23. Baumel A, Yom-Tov E. Predicting user adherence to behavioral eHealth interventions in the real world: examining which aspects of intervention design matter most. Transl Behav Med 2018;8(5):793-798. [doi: 10.1093/tbm/ibx037] [Medline: 29471424]

24. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]

25. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive health communication applications for people with chronic disease. Cochrane Database Syst Rev 2005(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]

26. Muñoz RF. Using evidence-based internet interventions to reduce health disparities worldwide. J Med Internet Res 2010;12(5):e60 [FREE Full text] [doi: 10.2196/jmir.1463] [Medline: 21169162]

27. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]

28. Stellefson M, Chaney B, Chaney D. The digital divide in health education. Am J Health Educ 2008;39(2):106-112. [doi: 10.1080/19325037.2008.10599023]

29. Ofcom. 2018. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0018/124254/bbc-barb-analysis.pdf [accessed 2020-10-03]

30. Schramm H, Cohen EL. Emotion regulation and coping via media use. In: Rössler P, Hoffner CA, van Zoonen L, editors. The International Encyclopedia of Media Effects. Hoboken, NJ: John Wiley & Sons; 2017:1-9.

31. Stamatakis E, Hillsdon M, Mishra G, Hamer M, Marmot M. Television viewing and other screen-based entertainment in relation to multiple socioeconomic status indicators and area deprivation: the Scottish Health Survey 2003. J Epidemiol Community Health 2009;63(9):734-740. [doi: 10.1136/jech.2008.085902] [Medline: 19401278]

32. Alsubaie M, Abbott R, Dunn B, Dickens C, Keil TF, Henley W, et al. Mechanisms of action in mindfulness-based cognitive therapy (MBCT) and mindfulness-based stress reduction (MBSR) in people with physical and/or psychological conditions: a systematic review. Clin Psychol Rev 2017;55:74-91. [doi: 10.1016/j.cpr.2017.04.008] [Medline: 28501707]

33. Brijnath B, Protheroe J, Mahtani KR, Antoniades J. Do web-based mental health literacy interventions improve the mental health literacy of adult consumers? Results from a systematic review. J Med Internet Res 2016;18(6):e165 [FREE Full text] [doi: 10.2196/jmir.5463] [Medline: 27323907]

34. Fava GA. Well-Being Therapy: Treatment Manual and Clinical Applications. Bologna/Buffalo, NY: Karger; 2016.

35. Fennell M, Bennett-Levy J, Westbrook D. Depression. In: Bennett-Levy J, Butler G, Fennell M, Hackman A, Mueller M, Wesrbrook D, editors. Oxford Guide to Behavioural Experiments in Cognitive Therapy. Oxford, UK: Oxford University Press; 2004:205-222.

36. Goldfried MR. Obtaining consensus in psychotherapy: what holds us back? Am Psychol 2019;74(4):484-496. [doi: 10.1037/amp0000365] [Medline: 30221947]

37. McAleavey AA, Castonguay LG, Goldfried MR. Clinical experiences in conducting cognitive-behavioral therapy for social phobia. Behav Ther 2014;45(1):21-35. [doi: 10.1016/j.beth.2013.09.008] [Medline: 24411111]

38. Mehta N, Clement S, Marcus E, Stona AC, Bezborodovs N, Evans-Lacko S, et al. Evidence for effective interventions to reduce mental health-related stigma and discrimination in the medium and long term: systematic review. Br J Psychiatry 2015;207(5):377-384 [FREE Full text] [doi: 10.1192/bjp.bp.114.151944] [Medline: 26527664]

39. de Souza Tursi MF, von Werne Baes C, de Barros Camacho FR, de Carvalho Tofoli SM, Juruena MF. Effectiveness of psychoeducation for depression: a systematic review. Aust N Z J Psychiatry 2013;47(11):1019-1031. [doi: 10.1177/0004867413491154] [Medline: 23739312]

40. Van Daele T, Hermans D, Van Audenhove C, Van den Bergh O. Stress reduction through psychoeducation: a meta- analytic review. Health Educ Behav 2012;39(4):474-485. [doi: 10.1177/1090198111419202] [Medline: 21986242]

41. Singhal A, Rogers EM. A theoretical agenda for entertainment–education. Commun Theory 2002;12(2):117-135. [doi: 10.1111/j.1468-2885.2002.tb00262.x]

42. Doonan S. Drag: The Complete Story. London, UK: Laurence King Publishing; 2019.

43. Ofcom's EPG prominence consultation. ITV PLC's response. Ofcom. 2019. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0020/131186/ITV.pdf [accessed 2020-11-30]

44. Dryden W. Cognitive Behaviour Therapies. London, UK: Sage Publications; 2012.

45. Gilbert P. Compassion Focused Therapy: Distinctive Features. Hove, UK: Routledge; 2010.

46. Segal ZV, Williams JM, Teasdale JD. Mindfulness-Based Cognitive Therapy for Depression. New York, NY: Guilford Publications; 2018.

47. Martell CR, Dimidjian S, Herman-Dunn R. Behavioral Activation for Depression: A Clinician's Guide. New York, NY: Guilford Press; 2013.

48. Surveys using our four personal well-being questions: a guide to what surveys include the four ONS personal well-being questions. Office for National Statistics. 2018. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/methodologies/surveysusingthe4officefornationalstatisticspersonalwellbeingquestions [accessed 2020-05-14]

49. SOC2010 volume 1: structure and descriptions of unit groups. Office for National Statistics. 2010. URL: https://tinyurl.com/2p8namb4 [accessed 2020-06-20]

50. The National Statistics Socio-economic classification (NS-SEC). Office for National Statistics. 2016. URL: https://tinyurl.com/3rwfz7yy [accessed 2020-06-20]

51. McDonald M, Dunbar I. Socio-economic classification (United Kingdom). Market Segmentation. 2004. URL: https://www.marketsegmentation.co.uk/downloads/Socio-economic%20Classification%20-%20United%20Kingdom.pdf [accessed 2020-03-14]

52. Mason J. Qualitative Researching. 3rd Edition. London, UK: Sage Publications; 2017.

53. Data protection act 2018. United Kingdom Legislation. 2018. URL: https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted [accessed 2020-05-23]

54. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]

55. Turnbull S, Lucas PJ, Hay AD, Cabral C. The role of economic, educational and social resources in supporting the use of digital health technologies by people with T2D: a qualitative study. BMC Public Health 2021;21(1):293 [FREE Full text] [doi: 10.1186/s12889-021-10325-7] [Medline: 33546661]

56. Turnbull S. The influence of digital self-care interventions on health inequality in high burden chronic health conditions. University of Bristol. 2019. URL: https://research-information.bris.ac.uk/ws/portalfiles/portal/198133744/Final_Copy_2019_05_21_Turnbull_S_PhD_Redacted.pdf [accessed 2020-11-30]

57. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]

58. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]

59. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]

60. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]

61. Alqahtani F, Al Khalifah G, Oyebode O, Orji R. Apps for mental health: an evaluation of behavior change strategies and recommendations for future development. Front Artif Intell 2019;2:30 [FREE Full text] [doi: 10.3389/frai.2019.00030] [Medline: 33733119]

62. Singhal A, Cody MJ, Rogers EM, Sabido M. Entertainment-Education and Social Change: History, Research, and Practice. Mahwah, NJ: Lawrence Erlbaum Associates; 2004.

63. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]

64. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]

65. Senecal C, Widmer RJ, Bailey K, Lerman LO, Lerman A. Usage of a digital health workplace intervention based on socioeconomic environment and race: retrospective secondary cross-sectional study. J Med Internet Res 2018;20(4):e145 [FREE Full text] [doi: 10.2196/jmir.8819] [Medline: 29685862]

66. Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, et al. Using science to sell apps: evaluation of mental health app store quality claims. NPJ Digit Med 2019;2:18 [FREE Full text] [doi: 10.1038/s41746-019-0093-1] [Medline: 31304366]

67. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]

68. Coupe N, Cotterill S, Peters S. Tailoring lifestyle interventions to low socio-economic populations: a qualitative study. BMC Public Health 2018;18(1):967 [FREE Full text] [doi: 10.1186/s12889-018-5877-8] [Medline: 30075716]

69. Schueller SM, Neary M, O'Loughlin K, Adkins EC. Discovery of and interest in health apps among those with mental health needs: survey and focus group study. J Med Internet Res 2018;20(6):e10141 [FREE Full text] [doi: 10.2196/10141] [Medline: 29891468]

70. Rubin DB. Inference and missing data. Biometrika 1976;63(3):581-592. [doi: 10.2307/2335739]

71. Bethlehem J. Selection bias in web surveys. Int Stat Rev 2010;78(2):161-188. [doi: 10.1111/j.1751-5823.2010.00112.x]

Abbreviations

BCT: behavior change technique
NS-SEC: National Statistics Socio-economic Classification
ONS: Office for National Statistics
OR: odds ratio

XSL•FO

RenderX

Original Paper

The Impact of Telemedicine Visits on the Controlling High Blood Pressure Quality Measure During the COVID-19 Pandemic: Retrospective Cohort Study

Siqin Ye[1,2], MSc, MD; D Edmund Anstey[1], MPH, MD; Anne Grauer[1], MD; Gil Metser[1], MD; Nathalie Moise[1,2], MSc, MD; Joseph Schwartz[2,3], PhD; Ian Kronish[1,2], MPH, MD; Marwah Abdalla[1,2], MPH, MD

[1]Department of Medicine, Columbia University Irving Medical Center, New York, NY, United States
[2]Center for Behavioral Cardiovascular Health, Columbia University Irving Medical Center, New York, NY, United States
[3]Department of Psychiatry and Behavioral Sciences, Stony Brook University, Stony Brook, NY, United States

Corresponding Author:
Siqin Ye, MSc, MD
Department of Medicine
Columbia University Irving Medical Center
622 West 168th Street
PH 9-318
New York, NY, 10033
United States
Phone: 1 7737100114
Fax: 1 2123053931
Email: sy2357@cumc.columbia.edu

Abstract

Background: Telemedicine visit use vastly expanded during the COVID-19 pandemic, and this has had an uncertain impact on cardiovascular care quality.

Objective: We sought to examine the association between telemedicine visits and the failure to meet the *Controlling High Blood Pressure* (BP) quality measure from the Centers for Medicare & Medicaid Services.

Methods: This was a retrospective cohort study of 32,727 adult patients with hypertension who were seen in primary care and cardiology clinics at an urban, academic medical center from February to December 2020. The primary outcome was a failure to meet the *Controlling High Blood Pressure* quality measure, which was defined as having no BP recorded or having a last recorded BP of ≥140/90 mm Hg (ie, poor BP control). Multivariable logistic regression was used to assess the association between telemedicine visit use during the study period (none, 1 telemedicine visit, or ≥2 telemedicine visits) and poor BP control; we adjusted for demographic and clinical characteristics.

Results: During the study period, no BP was recorded for 2.3% (486/20,745) of patients with in-person visits only, 27.1% (1863/6878) of patients with 1 telemedicine visit, and 25% (1277/5104) of patients with ≥2 telemedicine visits. After adjustment, telemedicine use was associated with poor BP control (1 telemedicine visit: odds ratio [OR] 2.06, 95% CI 1.94-2.18; *P*<.001; ≥2 telemedicine visits: OR 2.49, 95% CI 2.31-2.68; *P*<.001; reference: in-person visits only). This effect disappeared when the analysis was restricted to patients with at least 1 recorded BP (1 telemedicine visit: OR 0.89, 95% CI 0.83-0.95; *P*=.001; ≥2 telemedicine visits: OR 0.91, 95% CI 0.83-0.99; *P*=.03).

Conclusions: Increased telemedicine visit use is associated with poorer performance on the *Controlling High Blood Pressure* quality measure. However, telemedicine visit use may not negatively impact BP control when BP is recorded.

Introduction

The COVID-19 pandemic resulted in the rapid expansion of telemedicine as an integral part of outpatient care [1-7]. A telemedicine visit is defined as using video and telephone technology to connect patients to their clinicians as a substitute for in-person office visits. Telemedicine visits have helped maintain the continuity of care and preserve patients' access to care and are expected to remain a significant part of the health care delivery landscape even after the COVID-19 pandemic ends [8,9]. However, despite its increased utilization, concern has arisen that the content of care delivered via telemedicine visits may differ from that delivered during in-person visits. In particular, studies have shown that telemedicine visits are less likely to address standard components of care, such as blood pressure (BP) assessment, laboratory testing, and medication prescriptions and test orderings [10,11].

BP measurement is a fundamental component of hypertension management [12,13], and there has been increasing concern that the increased use of telemedicine visits in primary care and cardiology outpatient settings may impact the accurate assessment of BP and, in turn, how well BP is controlled for patients with hypertension [14]. To address this potentially unintended consequence of telemedicine expansion, we performed an electronic health record (EHR)–based retrospective analysis of a diverse population of patients with hypertension receiving care at a large, urban, academic medical center across an 11-month period during the COVID-19 pandemic. Using the specification of the *Controlling High Blood Pressure* quality measure from the Centers for Medicare & Medicaid Services (CMS) that defines poor BP control as having no recorded BP measurements or having a last recorded BP of ≥140/90 mm Hg [15,16], we sought to determine the association between telemedicine visit use and the failure to meet this widely used measure for benchmarking population-level quality of care, hypothesizing that the increased use of telemedicine visits can lead to poorer performance on this quality measure.

Methods

Ethics Approval

The study was approved by the Columbia University Irving Medical Center Institutional Review Board (AAAT0375), in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. Informed consent was waived due to the study being of minimal risk.

Study Design

After approval by our institutional review board, we queried our EHR (Epic EHR; Epic Systems) to identify all completed outpatient visits to primary care and cardiology clinics at Columbia University Irving Medical Center from February 1, 2020, to December 31, 2020. The Medical Center transitioned to Epic EHR on February 1, 2020, that is, 1 month prior to the start of the COVID-19 pandemic in New York City. For each visit, we extracted information, including patients' date of birth, sex, race, and ethnicity; the primary payer; and the International Classification of Diseases, 10th Revision (ICD-10) codes associated with the visits. BP data associated with the visits were extracted from Epic flow sheets, and BP measurements were recorded in accordance with standard clinical protocols (ie, staff or clinicians recorded BP measurements in the office during in-person visits, and self-reported BP values that were measured at home during telemedicine visits at clinicians' discretion).

For our analysis, we defined the study population by using specifications published by the CMS for the 2020 version of the quality measure *HTN-2: Controlling High Blood Pressure* [16]. Specifically, we included all patients aged between 18 and 85 years with a diagnosis of hypertension, which was defined as having the ICD-10 code "E10" associated with any visit during the study period. Patients with end-stage renal disease, those with a history of kidney transplant, or those who were pregnant during the study period were excluded based on the ICD-10 codes for these conditions published by the CMS. For the primary outcome, following the above CMS specification, we defined a patient as one who failed to meet the *Controlling High Blood Pressure* quality measure if (1) BP was not recorded during any visit included in the study period or (2) the last BP recorded during the study period was ≥140/90 mm Hg.

For the primary exposure variable—telemedicine visit use—we defined telemedicine visits based on appointment type (ie, visits that were scheduled and conducted by using video or telephone technology) [2]. Because the distribution of telemedicine visits is right-skewed, we classified patients into the following three categories: those with in-person visits only, those with 1 telemedicine visit, and those with ≥2 telemedicine visits. To adjust for the total number of visits (including in-person and telemedicine visits), we similarly defined the following patient categories: patients with a total number of 1, 2, or ≥3 outpatient visits during the study period.

We used patients' self-reported race and ethnicity to classify patients into the following race and ethnicity categories: non-Hispanic White; non-Hispanic Black; Hispanic; Asian, Hawaiian, and Pacific Islander; and other, declined to answer, or unknown. Patient insurance was categorized as "commercial," "Medicare," or "Medicaid" by using the primary payer field associated with the last visit during the study period. Using a similar approach as the one described above for hypertension, we identified patients with a diagnosis of atherosclerotic cardiovascular disease (ASCVD) and diabetes mellitus through visit-associated ICD-10 codes.

Statistical Analysis

Descriptive statistics were calculated for the proportion of visits in which BP was recorded (ie, for each type of visit [in-person, video, or telephone visits]). Descriptive statistics were also calculated for demographic and clinical characteristics according to the categories of telemedicine use. A chi-square test, a 1-way ANOVA, and the Kruskal-Wallis test were used for categorical variables, normally distributed continuous variables, and nonnormally distributed continuous variables, respectively. A multivariable logistic regression model was used to determine if telemedicine use was associated with higher odds of poor BP

XSL·FO

RenderX

control; we adjusted for age, race and ethnicity, payer, and the presence of comorbidities. Because telemedicine visits frequently do not have recorded BP data, we applied the same model for the subgroup of patients with at least 1 recorded BP. As disparities in telemedicine visit use have been previously described for older patients and by race, ethnicity, and insurance status [2,4,5], we conducted additional subgroup analyses for patients aged ≥65 years, non-Hispanic Black patients, Hispanic patients, and patients with Medicaid. All analyses were carried out by using Stata statistical software (version 16; StataCorp LLC).

Results

For the study population, we identified 32,727 patients aged between 18 and 85 years who had at least 1 completed outpatient visit to primary care or cardiology clinics at Columbia University Irving Medical Center between February 1, 2020, to December 31, 2020, and who had a diagnosis of hypertension (but not end-stage renal disease, history of kidney transplant, or pregnancy). Of these patients, 20,745 (63.3%) had an in-person visit only, 6878 (21%) had 1 telemedicine visit, and 5104 (15.6%) had ≥2 telemedicine visits. Detailed baseline characteristics are described in Table 1. Specifically, patients with more telemedicine visit use were more likely to be female, be Hispanic or non-Hispanic Black, and have Medicaid insurance. They were also less likely to have ASCVD but more likely to have diabetes mellitus, and they had a higher total number of visits.

Of the total 87,309 visits across the study period, 59,409 were conducted in person, 14,982 were video visits, and 12,918 were telephone visits. BP was recorded for 93% (55,370/59,409) of in-person visits, 20% (3011/14,982) of video visits, and 9% (1187/12,918) of telephone visits (Figure 1).

Table 1. Demographic and clinical characteristics of patients with hypertension by telemedicine use.

Characteristic	All in-person visits (n=20,745)	1 telemedicine visit (n=6878)	≥2 telemedicine visits (n=5104)	P value
Age (years), mean (SD)	66.7 (11.8)	65.7 (11.9)	65.4 (12.1)	.08
Sex, n (%)				<.001
Male	10,489 (50.6)	3325 (48.3)	1868 (36.6)	
Female	10,256 (49.4)	3553 (51.7)	3236 (63.4)	
Race and ethnicity, n (%)				<.001
Non-Hispanic White	7492 (36.1)	1946 (28.3)	883 (17.3)	
Hispanic	3127 (15.1)	1810 (26.3)	2484 (48.7)	
Non-Hispanic Black	1518 (7.3)	600 (8.7)	537 (10.5)	
Asian, Hawaiian, and Pacific Islander	485 (2.3)	140 (2)	90 (1.8)	
Other, declined to answer, or unknown	8123 (39.2)	2382 (28.3)	1110 (21.8)	
Primary insurance, n (%)				<.001
Commercial	7169 (34.6)	2150 (31.3)	927 (18.2)	
Medicare	11,865 (57.2)	3826 (55.6)	2982 (58.4)	
Medicaid	1711 (8.3)	902 (13.1)	1195 (23.4)	
Comorbidities, n (%)				
Atherosclerotic cardiovascular disease	6375 (30.7)	1990 (28.9)	1257 (24.6)	<.001
Diabetes mellitus	3439 (16.6)	1532 (22.3)	1792 (35.1)	<.001
Total number of visits, median (IQR)	2 (1-3)	2 (1-4)	4 (3-7)	<.001
Blood pressure[a] (mm Hg), mean (SD)				
Systolic blood pressure	132.4 (16.3)	132.5 (17.0)	134.3 (17.9)	<.001
Diastolic blood pressure	77.5 (9.7)	77.3 (9.7)	78.0 (10.0)	.004
Hypertension control status, n (%)				<.001
No blood pressure measured	486 (2.3)	1863 (27.1)	1277 (25)	
Last recorded blood pressure of <140/90 mm Hg	13,374 (64.5)	3346 (48.7)	2384 (46.7)	
Last recorded blood pressure of ≥140/90 mm Hg	6885 (33.2)	1863 (27.1)	1443 (28.3)	

[a]Calculated from patients with at least 1 recorded BP.

XSL•FO

RenderX

Figure 1. Number of visits by modality (in-person visit, video visit, or telephone visit). For each modality, whether BP was or was not recorded during the visit is also indicated. BP: blood pressure.
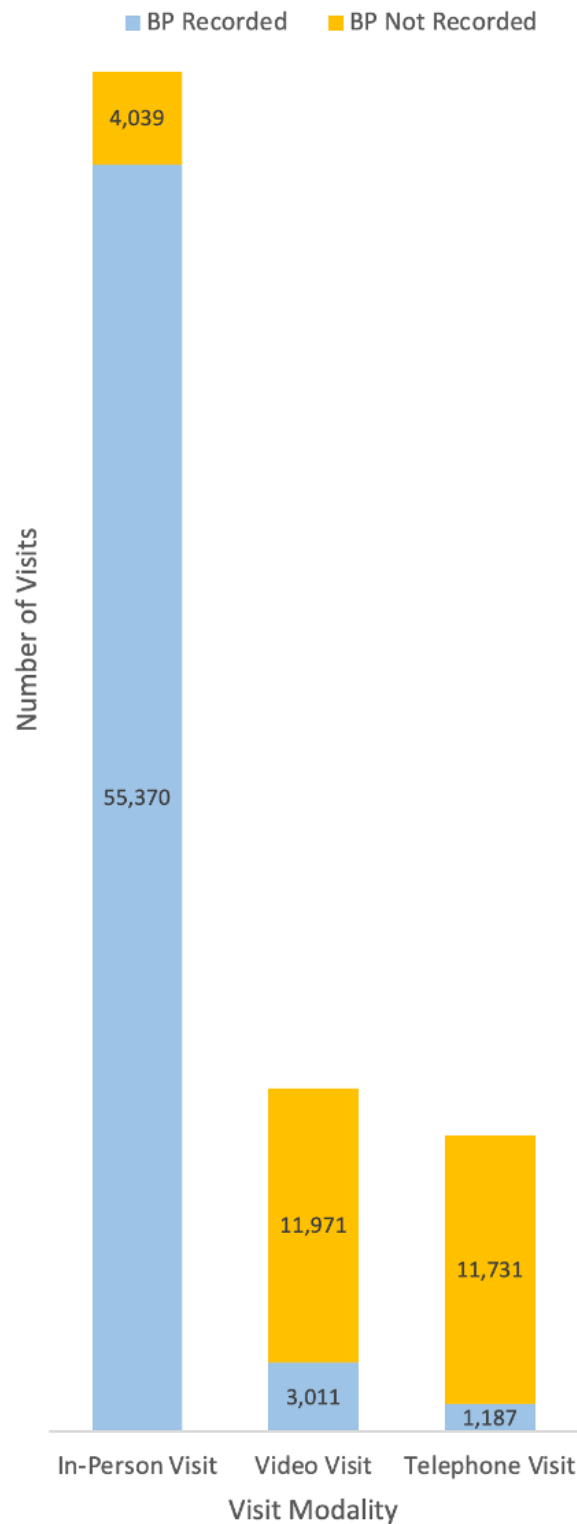
In a multivariable model that was adjusted for demographic and clinical characteristics, telemedicine use was associated with higher odds of not meeting the *Controlling High Blood Pressure* quality measure (1 telemedicine visit: odds ratio [OR] 2.06, 95% CI 1.94-2.18; *P*<.001; ≥2 telemedicine visits: OR 2.49, 95% CI 2.31-2.68; *P*<.001; reference: in-person visits only;

Table 2). Older age, Hispanic or Black patients, Medicaid insurance, and diabetes mellitus were also associated with higher odds of not meeting the measure, while ASCVD and 2 or ≥3 total visits during the study period were associated with lower odds of not meeting the measure.

Table 2. Multivariable analysis for predictors of the failure to meet the Controlling High Blood Pressure (BP) quality measure, including all patients with hypertension and only those with at least 1 recorded BP. The failure to meet the measure is defined as having (1) no BP recorded at any visit or (2) a last recorded BP of ≥140/90 mm Hg.

Predictors	All patients with hypertension (n=32,727)		Patients with at least 1 recorded BP (n=29,101)	
	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Age (per 10-year increase)	1.03 (1.00-1.05)	.04	1.02 (0.99-1.05)	.14
Sex				
Male	Reference	N/A[a]	Reference	N/A
Female	1.03 (0.99-1.08)	.18	1.07 (1.02-1.13)	.009
Race and ethnicity				
Non-Hispanic White	Reference	N/A	Reference	N/A
Hispanic	1.45 (1.36-1.55)	<.001	1.43 (1.33-1.54)	<.001
Non-Hispanic Black	1.58 (1.45-1.73)	<.001	1.63 (1.48-1.79)	<.001
Asian, Hawaiian, and Pacific Islander	0.97 (0.82-1.13)	.67	0.96 (0.81-1.14)	.65
Other, declined to answer, or unknown	1.03 (0.97-1.09)	.29	1.05 (0.99-1.12)	.10
Primary insurance				
Commercial	Reference	N/A	Reference	N/A
Medicare	0.97 (0.91-1.04)	.36	1.00 (0.94-1.08)	.94
Medicaid	1.28 (1.18-1.38)	<.001	1.28 (1.17-1.39)	<.001
Comorbidities				
Atherosclerotic cardiovascular disease	0.71 (0.68-0.75)	<.001	0.72 (0.68-0.76)	<.001
Diabetes mellitus	1.10 (1.04-1.16)	.002	1.12 (1.05-1.19)	<.001
Total number of visits				
1	Reference	N/A	Reference	N/A
2	0.72 (0.69-0.77)	<.001	1.09 (1.02-1.16)	.01
≥3	0.49 (0.46-0.52)	<.001	1.11 (1.04-1.19)	.002
Number of telemedicine visits				
In-person visit only	Reference	N/A	Reference	N/A
1 telemedicine visit	2.06 (1.94-2.18)	<.001	0.89 (0.83-0.95)	.001
≥2 telemedicine visits	2.49 (2.31-2.68)	<.001	0.91 (0.83-0.99)	.03

[a]N/A: not applicable.

When restricting the analysis to the 29,101 patients (patients with in-person visits only: n=20,259, 69.6%; patients with 1 telemedicine visit: n=5015, 17.2%; patients with ≥2 telemedicine visits: n=3825, 13.1%) with a least 1 recorded BP, telemedicine use was associated with lower odds of not meeting the *Controlling High Blood Pressure* quality measure (1 telemedicine visit: OR 0.89, 95% CI 0.83-0.95; *P*<.001; ≥2 telemedicine visits: OR 0.91, 95% CI 0.83-0.99; *P*=.03; reference, in-person visit only). In this model, female sex, Hispanic patients of color, Black patients, Medicaid insurance, and diabetes mellitus were associated with higher odds of not meeting the measure, while ASCVD and 2 or ≥3 total visits during the study period continued to be associated with lower odds of not meeting the measure (Table 2).

Subgroups analyses for patients who were aged ≥65 years, Hispanic patients, non-Hispanic Black patients, and those with Medicaid insurance are shown in Multimedia Appendix 1. The impact of telemedicine use on BP control in these subgroups was similar to that in the analysis that was restricted to the subgroup of patients with at least 1 recorded BP, both in the model that included all patients with hypertension and when only patients with at least 1 recorded BP were included.

Discussion

In our analysis of patients with hypertension who were seen in primary care and cardiology clinics at an urban, academic medical center during the COVID-19 pandemic in 2020, we found that increased telemedicine visit use was associated with poorer performance on the *Controlling High Blood Pressure* quality measure. These findings are largely driven by BP being recorded in less than 20% (4198/27,900) of telemedicine visits, as BP was recorded in 93% (55,370/59,409) of in-person visits.

When the analysis was restricted to patients with at least 1 recorded BP, patients with higher telemedicine visit use had a better or similar likelihood of poor performance on the *Controlling High Blood Pressure* quality measure. These findings were also robust for the subgroups of patients who were previously described to have more difficulty with using telemedicine services, including patients aged ≥65 years, Black or Hispanic patients, and those with Medicaid insurance [2,4,5,17,18].

Our finding that primary care and cardiology telemedicine visits are less likely to have recorded B*P* values when compared to in-person visits is consistent with prior literature. A recent report from a large US database containing 125.8 million primary care visits from 2018 to 2020 demonstrated that BP is recorded in less than 10% of telemedicine visits, whereas BP is recorded in approximately 70% of in-person visits [10]. In our analysis, we additionally found that the disparate rate of BP being recorded at telemedicine visits versus in-person visits resulted in patients who used telemedicine visits being more likely to not meet the *Controlling High Blood Pressure* quality measure. These findings highlight potential unintended consequence of the rapid adoption of telemedicine visits and have direct implications for various quality payment programs, such as those of Medicare accountable care organizations, that use BP control as a key quality benchmark [16]. More broadly, our findings highlight the importance of the continued assessment of the content and quality of care delivered via telemedicine services [10,11], especially as telemedicine visits are expected to remain an integral part of the health care delivery landscape after the COVID-19 pandemic ends.

Nonetheless, it is reassuring that telemedicine visit use did not negatively impact the *Controlling High Blood Pressure* quality measure when the analysis was restricted to patients with at least 1 recorded BP. This suggests that while BP is less likely to be recorded during telemedicine visits, in general the quality of BP management in primary care and cardiology settings may be similar regardless of telemedicine visit use. It is also reassuring that we observed similar results even in populations that are known to have decreased telemedicine use and poorer BP control, such as older patients, Black or Hispanic patients, or patients with Medicaid [2,4,5,19]. However, we cannot fully exclude residual confounding, such as from patients who use more telemedicine services also being more likely to maintain the continuity of care and other health behaviors during the COVID-19 pandemic [6,7,20]. Future studies would need to more rigorously evaluate how telemedicine use can impact BP management as well as determine the best approaches to incorporating BP management strategies, including BP telemonitoring, into routine clinical practice during the telemedicine era [21,22]. Examples of such strategies might include conducting randomized clinical trials of interventions to improve the accuracy of home BP assessments performed by patients in advance of telemedicine and office visits [23] as well as using approaches to integrating telemedicine visits as part of novel BP telemonitoring and medication titration programs [24,25].

There are several additional limitations to our study. As a retrospective cohort study that uses EHR data, our study is necessarily hypothesis-generating research, and additional confounding factors, including those outside of EHR data capture, cannot be excluded. Because we did not assess the quality of BP medications and potential clinical inertia, we have limited insight on how telemedicine use can affect clinical management practices for hypertension. Furthermore, the BP recordings used for this analysis reflect actual clinical practice, and we could not assess the quality of BP measurements that were taken at patents' homes and then recorded during telemedicine visits, although home BP measurements have been shown to be potentially more reliable than BPs measured during office visits [23,26]. The *Controlling High Blood Pressure* quality measure may also not accurately reflect a patient's true BP control status at a given point in time. Finally, because there were dramatic care disruptions during the COVID-19 pandemic, our findings may not be generalizable to in-person visits and telemedicine visits when the COVID-19 pandemic ends, and longitudinal research is needed to assess the continued impact of telemedicine visits on cardiovascular care delivery.

Despite these limitations, our study is among the first to describe the real-world impact of telemedicine visits on BP control for a diverse population of patients who access ambulatory care. We found that while the higher use of telemedicine visits was associated with poorer performance on the *Controlling High Blood Pressure* quality measure, this was mainly driven by BP being much less likely to be recorded during telemedicine visits. When BP was recorded, telemedicine use was found to be associated with similar or slightly improved BP control. These results provide timely insights into the impact of telemedicine on cardiovascular care quality with important implications for research, implementation, and policy making in the telemedicine era.

Conflicts of Interest

None declared.

XSL•FO

RenderX

Multimedia Appendix 1

Subgroup analyses on the association between telemedicine use and the failure to meet the "Controlling High Blood Pressure" quality measure for patient populations with potential technology barriers (age: ≥65 years; primary insurance: Medicaid; race and ethnicity: Hispanic people of color and non-Hispanic Black). Poor BP control is defined as having no BP recorded at any visit or having a last recorded BP of ≥140/90 mm Hg. BP: blood pressure; OR: odds ratio.

[DOCX File , 15 KB - formative_v6i3e32403_app1.docx]

References

1. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: Evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]

2. Ye S, Kronish I, Fleck E, Fleischut P, Homma S, Masini D, et al. Telemedicine expansion during the COVID-19 pandemic and the potential for technology-driven disparities. J Gen Intern Med 2021 Jan;36(1):256-258 [FREE Full text] [doi: 10.1007/s11606-020-06322-y] [Medline: 33105000]

3. Patel SY, Mehrotra A, Huskamp HA, Uscher-Pines L, Ganguli I, Barnett ML. Trends in outpatient care delivery and telemedicine during the COVID-19 pandemic in the US. JAMA Intern Med 2021 Mar 01;181(3):388-391 [FREE Full text] [doi: 10.1001/jamainternmed.2020.5928] [Medline: 33196765]

4. Eberly LA, Kallan MJ, Julien HM, Haynes N, Khatana SAM, Nathan AS, et al. Patient characteristics associated with telemedicine access for primary and specialty ambulatory care during the COVID-19 pandemic. JAMA Netw Open 2020 Dec 01;3(12):e2031640 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.31640] [Medline: 33372974]

5. Eberly LA, Khatana SAM, Nathan AS, Snider C, Julien HM, Deleener ME, et al. Telemedicine outpatient cardiovascular care during the COVID-19 pandemic: Bridging or opening the digital divide? Circulation 2020 Aug 04;142(5):510-512 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.120.048185] [Medline: 32510987]

6. Weiner JP, Bandeian S, Hatef E, Lans D, Liu A, Lemke KW. In-person and telehealth ambulatory contacts and costs in a large US insured cohort before and during the COVID-19 pandemic. JAMA Netw Open 2021 Mar 01;4(3):e212618 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.2618] [Medline: 33755167]

7. Whaley CM, Pera MF, Cantor J, Chang J, Velasco J, Hagg HK, et al. Changes in health services use among commercially insured US populations during the COVID-19 pandemic. JAMA Netw Open 2020 Nov 02;3(11):e2024984 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.24984] [Medline: 33151319]

8. Mehrotra A, Bhatia RS, Snoswell CL. Paying for telemedicine after the pandemic. JAMA 2021 Feb 02;325(5):431-432. [doi: 10.1001/jama.2020.25706] [Medline: 33528545]

9. Shachar C, Engel J, Elwyn G. Implications for telehealth in a postpandemic future: Regulatory and privacy issues. JAMA 2020 Jun 16;323(23):2375-2376. [doi: 10.1001/jama.2020.7943] [Medline: 32421170]

10. Alexander GC, Tajanlangit M, Heyward J, Mansour O, Qato DM, Stafford RS. Use and content of primary care office-based vs telemedicine care visits during the COVID-19 pandemic in the US. JAMA Netw Open 2020 Oct 01;3(10):e2021476 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.21476] [Medline: 33006622]

11. Yuan N, Pevnick JM, Botting PG, Elad Y, Miller SJ, Cheng S, et al. Patient use and clinical practice patterns of remote cardiology clinic visits in the era of COVID-19. JAMA Netw Open 2021 Apr 01;4(4):e214157 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.4157] [Medline: 33818619]

12. Muntner P, Shimbo D, Carey RM, Charleston JB, Gaillard T, Misra S, et al. Measurement of blood pressure in humans: A scientific statement from the American Heart Association. Hypertension 2019 May;73(5):e35-e66 [FREE Full text] [doi: 10.1161/HYP.0000000000000087] [Medline: 30827125]

13. Whelton PK, Carey RM, Aronow WS, Casey DEJ, Collins KJ, Himmelfarb CD, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension 2018 Jun;71(6):e13-e115 [FREE Full text] [doi: 10.1161/HYP.0000000000000065] [Medline: 29133356]

14. Bress AP, Cohen JB, Anstey DE, Conroy MB, Ferdinand KC, Fontil V, et al. Inequities in hypertension control in the United States exposed and exacerbated by COVID-19 and the role of home blood pressure and virtual health care during and after the COVID-19 pandemic. J Am Heart Assoc 2021 Jun;10(11):e020997 [FREE Full text] [doi: 10.1161/JAHA.121.020997] [Medline: 34006116]

15. Controlling high blood pressure (CBP). National Committee for Quality Assurance. URL: https://www.ncqa.org/hedis/measures/controlling-high-blood-pressure/ [accessed 2019-05-11]

16. HTN-2 (NQF 0018): Controlling high blood pressure. Centers for Medicare & Medicaid Services. 2020. URL: https://qpp.cms.gov/docs/QPP_quality_measure_specifications/Web-Interface-Measures/2020_Measure_HTN2_CMSWebInterface.pdf [accessed 2021-05-11]

17. Roberts ET, Mehrotra A. Assessment of disparities in digital access among Medicare beneficiaries and implications for telemedicine. JAMA Intern Med 2020 Oct 01;180(10):1386-1389 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2666] [Medline: 32744601]

18. Lam K, Lu AD, Shi Y, Covinsky KE. Assessing telemedicine unreadiness among older adults in the United States during the COVID-19 pandemic. JAMA Intern Med 2020 Oct 01;180(10):1389-1391 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2671] [Medline: 32744593]

19. Muntner P, Hardy ST, Fine LJ, Jaeger BC, Wozniak G, Levitan EB, et al. Trends in blood pressure control among US adults with hypertension, 1999-2000 to 2017-2018. JAMA 2020 Sep 22;324(12):1190-1200 [FREE Full text] [doi: 10.1001/jama.2020.14545] [Medline: 32902588]

20. Imlach F, McKinlay E, Middleton L, Kennedy J, Pledger M, Russell L, et al. Telehealth consultations in general practice during a pandemic lockdown: survey and interviews on patient experiences and preferences. BMC Fam Pract 2020 Dec 13;21(1):269 [FREE Full text] [doi: 10.1186/s12875-020-01336-1] [Medline: 33308161]

21. Timpel P, Oswald S, Schwarz PEH, Harst L. Mapping the evidence on the effectiveness of telemedicine interventions in diabetes, dyslipidemia, and hypertension: An umbrella review of systematic reviews and meta-analyses. J Med Internet Res 2020 Mar 18;22(3):e16791 [FREE Full text] [doi: 10.2196/16791] [Medline: 32186516]

22. Xu Y, Zhou X, Suehs BT, Hartzema AG, Kahn MG, Moride Y, et al. A comparative assessment of observational medical outcomes partnership and mini-sentinel common data models and analytics: Implications for active drug safety surveillance. Drug Saf 2015 Aug;38(8):749-765. [doi: 10.1007/s40264-015-0297-5] [Medline: 26055920]

23. Shimbo D, Artinian NT, Basile JN, Krakoff LR, Margolis KL, Rakotz MK, American Heart Association and the American Medical Association. Self-measured blood pressure monitoring at home: A joint policy statement from the American Heart Association and American Medical Association. Circulation 2020 Jul 28;142(4):e42-e63. [doi: 10.1161/CIR.0000000000000803] [Medline: 32567342]

24. Scirica BM, Cannon CP, Fisher NDL, Gaziano TA, Zelle D, Chaney K, et al. Digital care transformation: Interim report from the first 5000 patients enrolled in a remote algorithm-based cardiovascular risk management program to improve lipid and hypertension control. Circulation 2021 Feb 02;143(5):507-509 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.120.051913] [Medline: 33201729]

25. Xu H, Long H. The effect of smartphone app-based interventions for patients with hypertension: Systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Oct 19;8(10):e21759 [FREE Full text] [doi: 10.2196/21759] [Medline: 33074161]

26. Schwartz JE, Muntner P, Kronish IM, Burg MM, Pickering TG, Bigger JT, et al. Reliability of office, home, and ambulatory blood pressure measurements and correlation with left ventricular mass. J Am Coll Cardiol 2020 Dec 22;76(25):2911-2922 [FREE Full text] [doi: 10.1016/j.jacc.2020.10.039] [Medline: 33334418]

Abbreviations

ASCVD: atherosclerotic cardiovascular disease
BP: blood pressure
CMS: Centers for Medicare & Medicaid Services
EHR: electronic health record
ICD-10: International Classification of Diseases, 10th Revision
NHLBI: National Heart, Lung, and Blood Institute
OR: odds ratio

<u>Original Paper</u>

Identity Threats as a Reason for Resistance to Artificial Intelligence: Survey Study With Medical Students and Professionals

Ekaterina Jussupow[1*], PhD; Kai Spohrer[2*], PhD; Armin Heinzl[1*], Prof Dr

[1]University of Mannheim, Mannheim, Germany

[2]Frankfurt School of Finance & Management, Frankfurt, Germany

[*]all authors contributed equally

Corresponding Author:
Ekaterina Jussupow, PhD
University of Mannheim
L15 1-6
Mannheim, 68313
Germany
Phone: 49 621 181 1691
Email: jussupow@uni-mannheim.de

Abstract

Background: Information systems based on artificial intelligence (AI) have increasingly spurred controversies among medical professionals as they start to outperform medical experts in tasks that previously required complex human reasoning. Prior research in other contexts has shown that such a technological disruption can result in professional identity threats and provoke negative attitudes and resistance to using technology. However, little is known about how AI systems evoke professional identity threats in medical professionals and under which conditions they actually provoke negative attitudes and resistance.

Objective: The aim of this study is to investigate how medical professionals' resistance to AI can be understood because of professional identity threats and temporal perceptions of AI systems. It examines the following two dimensions of medical professional identity threat: threats to physicians' expert status (professional recognition) and threats to physicians' role as an autonomous care provider (professional capabilities). This paper assesses whether these professional identity threats predict resistance to AI systems and change in importance under the conditions of varying professional experience and varying perceived temporal relevance of AI systems.

Methods: We conducted 2 web-based surveys with 164 medical students and 42 experienced physicians across different specialties. The participants were provided with a vignette of a general medical AI system. We measured the experienced identity threats, resistance attitudes, and perceived temporal distance of AI. In a subsample, we collected additional data on the perceived identity enhancement to gain a better understanding of how the participants perceived the upcoming technological change as beyond a mere threat. Qualitative data were coded in a content analysis. Quantitative data were analyzed in regression analyses.

Results: Both threats to professional recognition and threats to professional capabilities contributed to perceived self-threat and resistance to AI. Self-threat was negatively associated with resistance. Threats to professional capabilities directly affected resistance to AI, whereas the effect of threats to professional recognition was fully mediated through self-threat. Medical students experienced stronger identity threats and resistance to AI than medical professionals. The temporal distance of AI changed the importance of professional identity threats. If AI systems were perceived as relevant only in the distant future, the effect of threats to professional capabilities was weaker, whereas the effect of threats to professional recognition was stronger. The effect of threats remained robust after including perceived identity enhancement. The results show that the distinct dimensions of medical professional identity are affected by the upcoming technological change through AI.

Conclusions: Our findings demonstrate that AI systems can be perceived as a threat to medical professional identity. Both threats to professional recognition and threats to professional capabilities contribute to resistance attitudes toward AI and need to be considered in the implementation of AI systems in clinical practice.

Introduction

Objective

Advances in machine learning and image recognition have driven the implementation of systems based on artificial intelligence (AI) in clinical practice. These systems are able to perform tasks commonly associated with intelligent beings, such as reasoning or learning from experience [1]. AI systems based on machine learning gain insight directly from large sets of data [2]. Such AI systems can, therefore, make more complex decisions and complete more complex tasks than can rule-based systems. Moreover, machine learning has the potential to improve AI system capabilities with growing amounts of training data. In medical disciplines such as radiology, AI systems already diagnose specific diseases in a manner comparable with that of experienced physicians [3]. Current AI systems provide diagnosis and treatment recommendations, facilitate access to information, and perform time-consuming routine tasks of physicians, such as the segmentation of radiological images [4]. In the light of an accelerated technological development, AI systems in health care are expected to pave the way for truly personalized medicine by augmenting medical decisions and helping physicians cope with increasing amounts of relevant data and growing numbers of medical guidelines [5]. In fact, future AI systems are expected to autonomously execute medical actions (eg, communicating results and making diagnosis decisions) as efficiently and accurately as expert physicians [3]. Thus, medical professionals will be able to delegate more complex medical tasks to those systems than to any traditional rule-based clinical decision support system [6].

Despite their benefits and the great expectations toward their future use, AI systems do not cause only positive reactions in medical professionals. Although nearly all current applications of AI are still limited to narrow use cases, the future potential of this technology has resulted in a discourse driven by a duality of hyped expectations and great fears [7]. Many physicians seem to be concerned about the changes that AI systems impose on their profession [8]. Such negative attitudes toward new technology can manifest as *resistance* attitudes, resulting in hesitation toward adopting a technology [9] and even in active resentment against using a technology in clinical practice [10]. Although multiple studies have investigated attitudes toward AI [11-13], they did not consider how resistance attitudes toward AI are formed. Specifically, they did not examine whether negative attitudes toward AI stem from the perception that AI is threatening the medical professional identity. Therefore, we aim to address the following research questions:

1. How is the medical professional identity threatened by AI systems?
2. Under which conditions do medical professional identity threats contribute to resistance to AI?

Theoretical Background

Professional identity refers to *how professionals define themselves in terms of their work roles and provides an answer to the question* "Who am I as a professional?" [14]. In general, professionals strive to maintain a coherent and positive picture of themselves [15], resulting in a tendency to interpret experiences in identity-coherent ways. This helps individuals to adapt to social changes, such as technological innovations, and to create benefits from the positive identification with a social group. Professional identity can be considered as a combination of social identity [16] (ie, membership in a group of professionals) and personal identity (ie, the individual identification and enactment of professional roles) [15]. Medical professionals are known for their strong commitment to their professional identity, which is already developed in the early phases of their socialization and later refined through practical experience [17]. The medical profession's group boundaries are very rigid and shaped by strong core values and ideals [18]. Medical professionals can, therefore, be seen as a *prototypical profession* [17], and their professional identity is particularly resilient to change [19].

Identity threats are "experiences appraised as indicating potential harm to the value, meanings, or enactment of an identity" [20]. Such experiences are potentially decreasing an identity's value, eliminating a part of the identity, or disturbing the connection between an identity and the meaning the individual associates with the threatened identity [20]. In the following, we distinguish between two parts of individuals' identity that can be threatened: personal identity and professional identity.

First, *self-threat* describes a context-independent threat to *personal identity* by challenging fundamental self-motives of distinctiveness, continuity, generalized self-efficacy, and self-esteem [21]. Self-threat can bias information processing, result in avoidance of threatening information [22-24], and adverse emotional reactions [25]. Recently, self-threat has been identified as an antecedent of resistance to technology [26].

Second, medical professionals can be threatened along different dimensions of their *professional identity*. Drawing on a synthesis of prior work (Multimedia Appendix 1 [9,10,14,27-34]), we differentiate between 2 dimensions of medical professional identity threats. First, *threats to professional recognition* refer to challenges to the expertise and status position of medical professionals [10,27]. Second, *threats to professional capabilities* refer to the enactment of roles related to the medical work itself. The latter include threats to the care provider role [28], autonomy [14,29-31], and professional control [9,32]. Multiple studies show that professional identity threats can stipulate resistance to new technologies and organizational change [10,28,35]; however, it is unclear how these threats manifest in the context of AI systems.

Multiple conditions do influence how medical professionals experience identity threat from AI systems. In this paper, we focus on the following two conditions: professional experience and the perceived temporal distance of AI. First, perceived

experience influences how likely medical professionals perceive that AI systems can replace parts of their work. In particular, more experienced physicians believe that they have unique skills that AI systems cannot substitute, whereas novices have yet to develop those skills. Second, medical professionals might have different perceptions of how fast AI systems are implemented in clinical practice. Perceiving AI systems as temporally close and relevant in the immediate future suggests that AI systems will be seen as more influential on concrete medical work practices, thus threatening one's professional capabilities. Conversely, if AI systems are perceived to be relevant only in the distant future, the perceived threat might be less specific to medical work practices but more relevant for the long-term reputation of medical professionals, thus threatening their professional recognition. Hence, the perceived temporal distance of AI systems could affect how relevant each dimension of the professional identity becomes [36].

Methods

Survey Design

We collected data in 2 waves of a web-based survey. The first wave survey mainly addressed medical students, whereas the second survey focused on experienced physicians from different specialties. All participants were provided with a vignette of an AI system named *Sherlock* (Textbox 1) that was based on the description of IBM Watson and was pretested with researchers and medical students. We selected this vignette because it depicts a general AI system in which the participants were familiar with because of the marketing efforts of the vendor. It does not limit the abilities of the AI to a specific medical specialty. The vignette was purposefully focused on the benefits of the system and evoked expectations of high accuracy to establish the picture of a powerful AI system that goes beyond extant narrow use cases and, thus, has the potential to be threatening. Then, control questions were included to ensure that participants associated AI with the vignette and that it was perceived as realistic. The vignette was pretested with medical students and professionals. We then asked an open question about the participants' perceptions of the changes to their professional role caused by AI systems to gain qualitative insights into the perceived upcoming change of their identity. Afterward, we asked participants to complete the provided survey of experience, identity threat, resistance attitudes, and perceived temporal distance of AI systems.

Textbox 1. Vignette Sherlock—a general artificial intelligence system.

What is an intelligent clinical decision support software?

- Physicians often need to quickly analyze all the information provided to make diagnoses and decisions about treatments. These decisions have far-reaching consequences for patients and yet often have to be made under time pressure and with incomplete information. This is where the Sherlock decision support software comes in.

- Sherlock can be used in different specialties but will be presented here as with the following example of an oncological system.

- Every physician has an iPad or a laptop through which he or she can access his or her electronic medical records. The "Ask Sherlock" button is integrated into each medical record. When the physician asks Sherlock, he or she receives a 1-page treatment recommendation with up to 40 pages of explanations as backup.

- With the help of artificial intelligence, Sherlock can integrate and compare millions of patient data to identify similarities and connections. In addition, Sherlock has been trained by renowned experts. On the basis of the evidence base and the expert training, Sherlock then generates treatment options. Sherlock then presents those treatment options ranked by appropriateness (recommended, for consideration, or not recommended) alongside key information from the medical record and relevant supporting articles from current research. As a result, the practicing physician can easily follow Sherlock's recommendations and directly access relevant articles or clinical data.

- Sherlock is already in use in some clinics and tests have shown that Sherlock's recommendations are 90% in line with the decisions of renowned tumor boards.

Measures

We used a 4-item measure of self-threat [21] on a 6-point Likert scale and a 3-item measure of resistance [9]. Furthermore, we asked participants whether they perceived the change from AI systems as temporally close or distant by assessing the agreement to the following statements: "Such systems will only become relevant in the distant future," "Such systems are unlikely to be implemented technically," and "Such systems are too abstract and intangible for me." We extended existing measures for threats toward professional recognition and professional capabilities following the procedure of MacKenzie et al [37], as outlined in Multimedia Appendix 2 [9,28,29,37-46]. For medical students, we also assessed positive expectations toward AI, which mirrored the negatively framed items that we used for identity threat (*identity enhancement*). We also included an open question about their expectations of how the medical role would change with the introduction of AI systems. As a control variable, we asked for participants' familiarity with clinical decision support systems. Except items related to self-threat, all items were measured on a 5-point Likert scale from totally disagree to totally agree. Table 1 lists the survey items. The items for *identity enhancement* can be found in Multimedia Appendix 3 [47-50].

XSL•FO

RenderX

Table 1. Final list of items used for the hypothesis testing.[a]

Construct	Item

Threats to professional recognition (self-developed from literature review)

Threat to expertise

E2: I fear that when using the system, physicians may lose their expert status.

E3: I fear that when using the system, certain physician specializations can be replaced.

Perceived threat to status position

S1: I fear that when using the system, physicians' position in the hospital hierarchy may be undermined.

S2: I fear that when using the system, physicians may have a lower professional status.

S4: I fear that the status of physicians, who use the system, may deteriorate within the physician community.

Threats to professional capabilities

Perceived threat to autonomy (adapted from Walter and Lopez [29])

A1: I fear that when using the system physicians' job autonomy may be reduced.

A3: I fear that physicians' diagnostic and therapeutic decisions will be more monitored by nonphysicians.

Perceived threat to professional influence

I1: I fear that when using the system physicians may have less control over patient medical decisions.

I2: I fear that when using the system physicians may have less control over ordering patient tests.

I3: I fear that when using the system physicians may have less control over the distribution of scarce resources.

Perceived threat to being a care provider (self-developed from literature review)

C1: I fear that when using the system physicians have less influence on patient care.

C3: I fear that when using the system physicians are less able to treat their patients well.

Self-threat from AI[b] (adapted from Murtagh et al [21])

ST1: Using Sherlock undermines my sense of self-worth.

ST2: Using Sherlock makes me feel less competent.

ST3: Using Sherlock would have to change who I am.

ST4: Using Sherlock makes me feel less unique as a person.

Resistance to AI (adapted from Bhattacherjee and Hikmet [9])

RC1: I do not want Sherlock to change the way I order patient tests.

RC2: I do not want Sherlock to change the way I make clinical decisions.

RC3: I do not want Sherlock to change the way I interact with other people on my job.

RC4: Overall, I do not want Sherlock to change the way I currently work.

Temporal distance of AI (self-developed)

A1: Such systems will only become relevant in the distant future.

A2: Such systems are unlikely be implemented technically.

A3: Such systems are too abstract and intangible for me.

Familiarity

F1: I have never heard of such systems to I have heard a lot of such systems

F2: I have never used such systems to I have used such systems quite often

F3: I have never dealt with such systems to I have dealt with such systems in great detail.

F4: I am not at all familiar with such systems to I am very familiar with such systems.

[a]Items with the identifiers E1, S3, A2, and C2 were removed because of measurement properties (see Multimedia Appendix 2).

[b]AI: artificial intelligence.

Scale Validation

We validated the scales of professional identity threats in the sample of novices and experienced physicians and the corresponding identity enhancement values, as outlined in Multimedia Appendix 2. A confirmatory factor analysis with all measurement scales resulted in a good model fit. All scales displayed good psychometric properties, including reliability, convergent validity, and discriminant validity (Multimedia Appendix 2). The correlation between threats to professional recognition and professional capabilities was 0.64, which was smaller than the lowest square root of the average variance extracted of 0.77, indicating acceptable multicollinearity. Similarly, multicollinearity between self-threat and threats to professional recognition was acceptable with a correlation of 0.66 being lower than the lowest square root of the average variance extracted of self-threat. We accounted for potential common method bias in the survey design and through testing for a common method factor. The results indicated that common method bias is unlikely to have a strong impact on our results (see Multimedia Appendix 2 for details). The items of identity enhancement were combined into 1 factor because of the result of the exploratory factor analysis. All analyses were performed using SPSS (version 26; IBM Corporation) and Stata (version 16; StataCorp).

Sample and Participants

A total sample of 227 novice and experienced physicians participated between fall 2017 and spring 2019. Participants were recruited from medical social media groups and by personal reference. After excluding participants because of failed comprehension checks or poor data quality (ie, very fast completion time or answers to the open question that were completely unrelated to the question), a total data sample of 206 participants was used for data analysis. Of these 206 participants, 164 (79.6%) were medical students and 42 (20.4%) participants were medical professionals across different specialties (see Table 2 for details on sample). We included both medical students and trained physicians in our sample for two reasons: First, especially inexperienced physicians are susceptible to the influence of technology [51]. They may thus provide valuable insight into the effects of AI systems. Second, particularly medical students face strong, long-term career consequences if AI systems alter the meaning of specific medical disciplines such as radiology. They are thus likely to cognitively engage with potential identity threats to make reasonable career decisions, for example, with regard to the specialty they pursue. Conversely, experienced medical professionals may have a more pronounced professional identity and may experience threats from AI systems differently.

Table 2. Sample properties of novice and experienced physicians.[a]

	Novice physicians	Experienced physicians
Total sample size, N	182	45
Sample used for data analysis, n (%)	164 (90.1)	42 (93.3)
Age (years), mean (SD)	24.65 (3.23)	39.57 (13.14)
Gender (female), n (%)	131 (72)	18 (40)
Experience (average)	Eighth semester	12.6 years of job experience

[a]The specialties of experienced physicians are presented at a later stage.

Statistical Analysis

Our main outcome variable measures participants' resistance toward using the *Sherlock* application. Because of differences in sample size between novice physicians (n=164) and experienced physicians (n=42), we conducted a nonparametric Mann–Whitney *U* test to test the differences between the 2 samples regarding resistance attitude and self-threat. We conducted seemingly unrelated regression analyses with self-threat and resistance as the dependent variables because we expected their error terms to be correlated. The predictors were included in a stepwise approach to assess how much additional variance they explain. To test the effect of perceived temporal distance of AI, we incorporated 2 interaction terms between temporal distance and the 2 dimensions of professional identity threats into the regression analysis. We conducted a mediation analysis following Hayes [52] with 10,000 bootstrap samples to test the mediating role of self-threat on resistance attitudes.

Ethics Approval

Ethics approval was not sought or obtained from the Ethics Committee at the University of Mannheim, as an approval was not required under the institution's current ethics statute.

Results

Descriptives and Group Differences

We find that both experienced and novice physicians perceived identity threats from the upcoming change from AI systems (Table 3). Novice physicians showed relatively high resistance to AI and self-threat, whereas experienced physicians showed slightly lower resistance and self-threat from AI. The group differences were significant for resistance (P=.005) and self-threat (P<.001). Novices perceived equally strong threats to their professional recognition (mean 3.05, SD 1.23) and professional capabilities (mean 3.25, SD 0.97), whereas experienced physicians perceived a stronger threat to their professional capabilities (mean 2.72, SD 1.18) than to their professional recognition (mean 2.38, SD 1.03). Group differences were statistically significant; novices experienced more threats to professional recognition (P<.001) and

professional capabilities (*P*=.80) than experienced physicians. Similarly, novices reported AI systems as slightly more temporally distant than did experienced physicians (*P*=.08). Moreover, in the sample of experienced physicians, the experienced threats and resistance attitudes differed based on the medical specialty (Table 4). The descriptive statistics show, for example, that physicians in the psychiatry specialty reported stronger threats to professional recognition than to professional capabilities, whereas surgeons reported stronger threats to professional capabilities than to professional recognition. However, as the sample size was relatively small, more research is needed to fully understand the effects of different specialties on experienced identity threats.

Table 3. Mann–Whitney *U* test between novice and experienced physicians.[a]

	Novice physicians (n=164), mean (SD)	Experienced physicians (n=42), mean (SD)	Group differences	
			Z value	*P* value
Resistance to AI[b]	3.58 (0.97)	3.13 (0.97)	2.83	.005
Self-threat from AI	2.78 (1.36)	1.92 (0.91)	3.85	<.001
Threats to professional recognition	3.05 (1.23)	2.38 (1.03)	3.36	.001
Threats to professional capabilities	3.25 (0.97)	2.72 (1.18)	2.62	.01
Perceived temporal distance of AI	2.23 (0.84)	2.03 (0.94)	1.75	.08
Familiarity with AI	1.75 (0.89)	2.32 (0.77)	−4.28	<.001

[a]All items except self-threat were measured on a 5-point Likert scale and self-threat was measured on a 6-point scale ranging from strongly disagree to strongly agree.

[b]AI: artificial intelligence.

Table 4. Means (SDs) by specialty of experienced physicians (n=42).

Specialty	Values, mean (SD)				
	Self-threat from AI[a]	Resistance to AI	Temporal distance of AI	Threats to professional recognition from AI	Threats to professional capabilities from AI
Not specified (n=7)	2.32 (0.89)	3.71 (1.25)	1.71 (0.65)	2.48 (0.87)	3.18 (1.33)
Internal medicine (n=10)	2.05 (0.98)	3.15 (0.88)	1.83 (0.65)	2.28 (1.33)	2.62 (1.27)
General medicine (n=3)	1.92 (1.01)	2.92 (0.88)	2.78 (1.68)	3.06 (0.59)	3.57 (0.95)
Psychiatry (n=5)	1.55 (0.82)	2.70 (0.97)	1.53 (0.61)	2.50 (0.90)	1.94 (0.73)
Pediatrics (n=5)	2.05 (1.25)	2.80 (1.14)	2.13 (1.17)	2.97 (0.84)	3.00 (1.06)
Surgery (n=5)	1.45 (0.45)	3.10 (1.10)	2.07 (0.64)	1.53 (0.63)	2.11 (1.44)
Anesthesiology (n=3)	1.75 (0.90)	3.58 (0.63)	1.78 (0.38)	2.94 (1.08)	2.96 (0.83)
Others such as neurology and pathology (n=4)	1.94 (1.13)	2.81 (0.59)	3.08 (1.52)	2.40 (0.83)	2.96 (1.04)

[a]AI: artificial intelligence.

Regression Analyses

Testing the relationships between different types of identity threat and resistance attitudes in the total sample (Tables 5 and 6; Multimedia Appendix 3), we found that perceived professional identity threats directly affected resistance attitudes and personal identity threat (self-threat). Both threats to professional recognition (*P*<.001) and threats to professional capabilities (*P*<.001) were significant predictors of self-threat (model 4a, Table 6). Moreover, we found that both professional identity threats contributed independently to resistance to change. However, threats to professional recognition predicted resistance only in isolation (Multimedia Appendix 2; *P*<.001) but not in combination with threats to professional capabilities (Multimedia Appendix 3, model 3b; *P*=.50). Hence, threats to professional capabilities overruled the impact of threats to professional recognition on resistance attitudes and significantly increased resistance (Multimedia Appendix 3, model 3b; *P*<.001). The findings suggest that threats to professional recognition are more strongly related to personal identity, whereas threats to professional capabilities are more strongly and directly related to resistance to change.

Table 5. Descriptive statistics of the measurement model (N=206).[a]

Variables	Mean (SD)	Square root of the AVE[b]	Resistance	Self-threat	Age	Gender	Gender	Temporal distance	ProCap[c]	ProRec[d]
Resistance	3.49 (0.989)	0.778	(0.856)	__e	—	—	—	—	—	—
Self-threat	2.606 (1.324)	0.821	0.491[f]	(0.891)	—	—	—	—	—	—
Age	27.689 (8.896)	—	–0.131[g]	–0.287[f]	(—)	—	—	—	—	—
Gender	0.345 (0.476)	—	–0.073	0.058	0.107	(—)	—	—	—	—
Familiarity	1.869 (0.892)	0.841	–0.122[g]	–0.152[h]	0.162[h]	0.124[g]	(0.905)	—	—	—
Temporal distance	2.188 (0.864)	0.673	0.264[f]	0.306[f]	–0.111	0.111	–0.209[f]	(0.713)	—	—
ProCap	3.14 (1.037)	0.771	0.529[f]	0.621[f]	–0.132[g]	0.006	–0.054	0.295[f]	(0.910)	—
ProRec	2.915 (1.133)	0.798	0.383[f]	0.660[f]	–0.154[h]	–0.050	–0.108	0.232[f]	0.640[f]	(0.896)

[a]Values in table are correlations between two variables. Values in parentheses are composite reliabilities.

[b]AVE: average variance extracted.

[c]ProCap: threats to professional capabilities.

[d]ProRec: threats to professional recognition.

[e]Not applicable.

[f]Significance level: $P<.001$.

[g]Significance level: $P<.05$.

[h]Significance level: $P<.01$.

Table 6. Results of seemingly unrelated hierarchical regression analyses with self-threat and resistance to change as dependent variables (full model 4).

|	Coefficient (SE; 95% CI)	Z value	P value
Model 4a with dependent variable self-threat			
Stage 1 (controls)			
Age	–0.034 (0.009; –0.052 to –0.016)	–3.650	<.001
Gender	–0.134 (0.134; –0.396 to 0.129)	–1.000	.39
Familiarity	–0.086 (0.072; –0.226 to 0.054)	–1.210	.25
Group (experienced and novice)	0.317 (0.218; –0.109 to 0.744)	1.460	.15
Step 2 (identity threats)			
ProRec[a]	0.503 (0.070; 0.366 to 0.641)	7.170	<.001
ProCap[b]	0.372 (0.080; 0.215 to 0.529)	4.650	<.001
Step 3 (Temporal distance of AI[c], interactions)			
Temporal distance	0.137 (0.077; –0.014 to 0.289)	1.780	.08
Temporal distance x ProRec	0.291 (0.078; 0.138 to 0.443)	3.730	<.001
Temporal distance x ProCap	–0.203 (0.086; –0.372 to –0.034)	–2.350	.02
Intercept	1.071 (0.276; 0.531 to 1.611)	3.880	<.001
Model 4b with dependent variable resistance			
Stage 1 (controls)			
Age	–0.001 (0.009; –0.018 to 0.016)	–0.130	.90
Gender	–0.134 (0.127; –0.382 to 0.114)	–1.060	.29
Familiarity	–0.066 (0.068; –0.198 to 0.067)	–0.970	.33
Group (experienced and novice)	–0.061 (0.206; –0.465 to 0.343)	–0.300	.77
Step 2 (identity threats)			
ProRec	0.055 (0.066; –0.076 to 0.185)	0.820	.41
ProCap	0.400 (0.076; 0.251 to 0.548)	5.270	<.001
Step 3 (Temporal distance of AI, interactions)			
Temporal distance	0.149 (0.073; 0.005 to 0.292)	2.030	.04
Temporal distance x ProRec	0.087 (0.074; –0.057 to 0.232)	1.190	.24
Temporal distance x ProCap	–0.165 (0.082; –0.325 to –0.005)	–2.020	.04
Intercept	0.237 (0.261; –0.274 to 0.748)	0.910	.36

[a]ProRec: threats to professional recognition.

[b]ProCap: threats to professional capabilities.
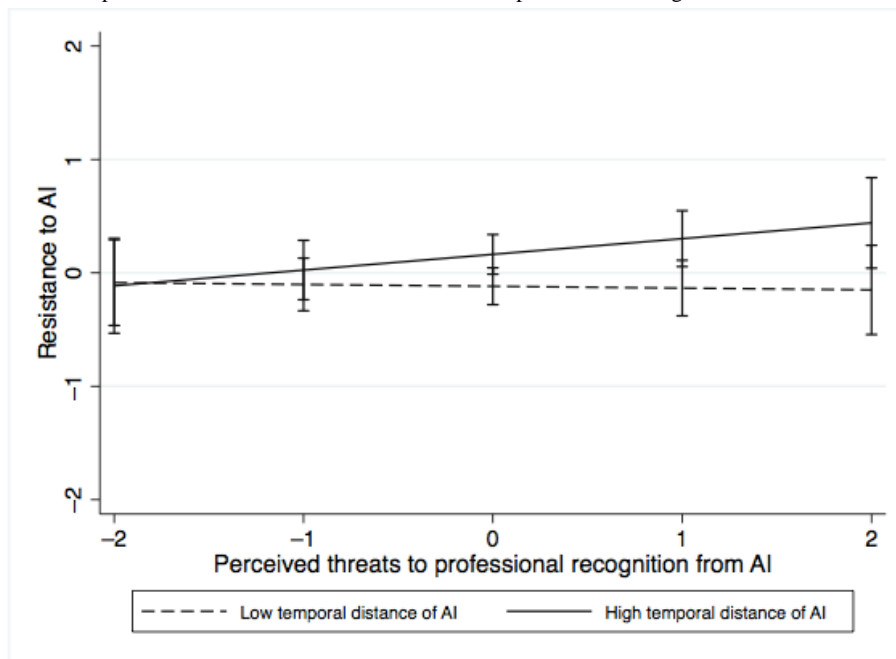
[c]AI: artificial intelligence.

We also analyzed how perceiving AI systems as temporally close or distant interacted with perceived professional identity threat. The perception of AI systems as temporally distant interacted positively with threats to professional recognition (P<.001) and interacted negatively with threats to professional capabilities (P=.02) in predicting self-threat (Table 6, model 4a). In predicting resistance, the perception of AI systems as temporally distant interacted negatively with threats to professional capabilities (Table 6, model 4b; P=.04), whereas the interaction with threats to professional recognition was not significant (P=.24). Figures 1-4 show the moderating effects of temporal distance on both dimensions of identity threat. The findings suggest that experienced identity threats are closely related to how temporally distant or close the technological change from AI systems is perceived. Threats to professional capabilities refer to more concrete and context-specific elements of professional identity. Thus, these threats are more salient if physicians believe that AI systems are temporally close and relevant to clinical practice in the near future. Conversely, threats to professional recognition require physicians to consider their profession in a holistic way. Thus, these threats are more salient if physicians perceive AI systems to be relevant only in the distant future.
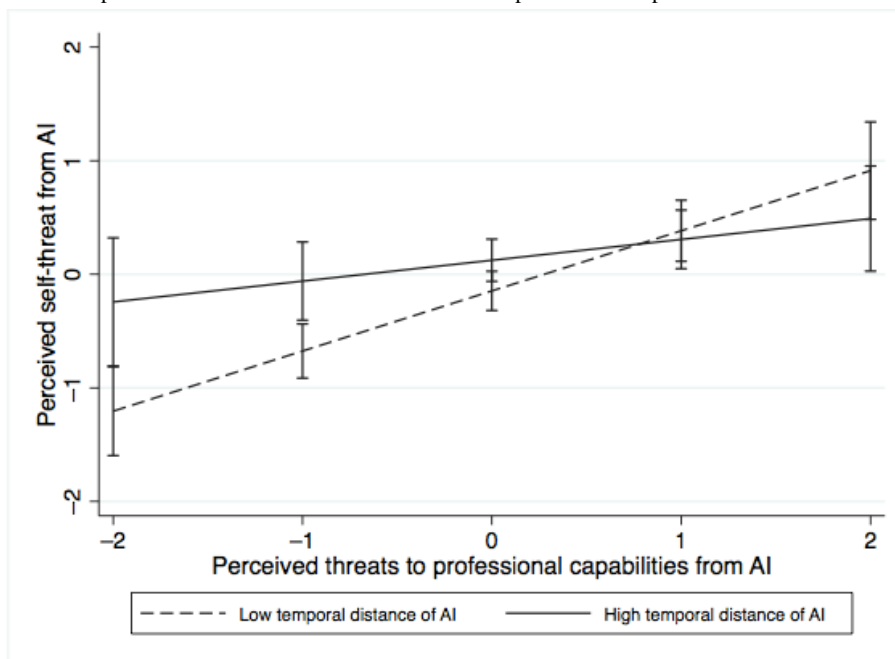
XSL•FO
RenderX

Figure 1. Moderating effect of temporal distance on the association of threats to professional recognition with self-threat. AI: artificial intelligence.
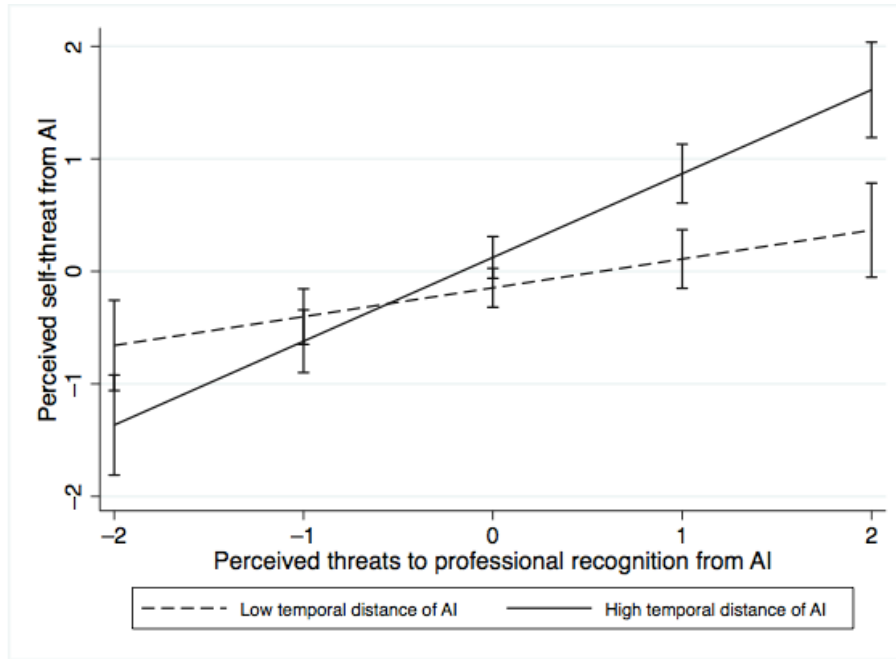

Figure 2. Moderating effect of temporal distance on the association of threats to professional recognition with resistance. AI: artificial intelligence.

Figure 3. Moderating effect of temporal distance on the association of threats to professional capabilities with self-threat. AI: artificial intelligence.

Figure 4. Moderating effect of temporal distance on the association of threats to professional capabilities with resistance. AI: artificial intelligence.
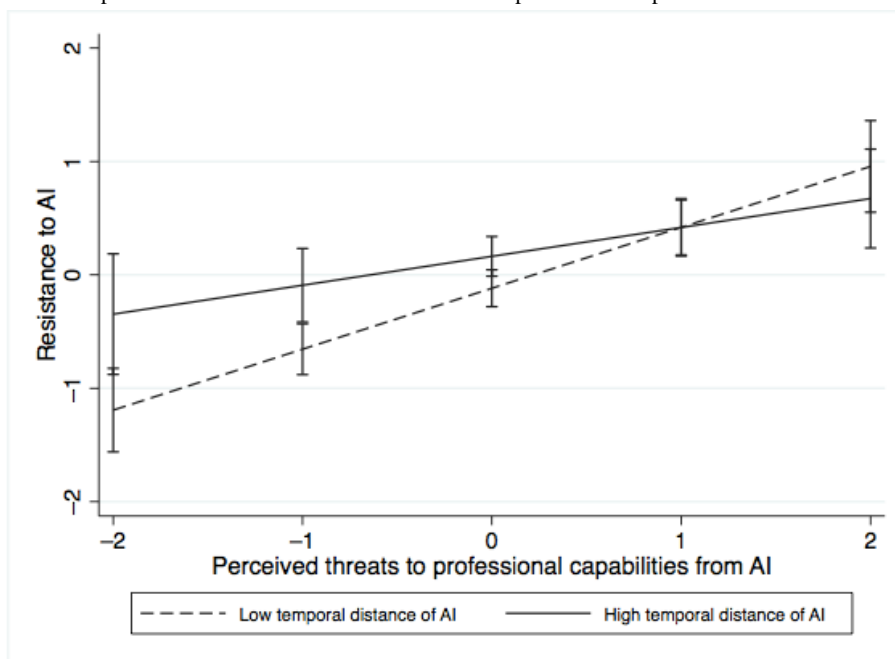
For the sample of novice physicians, we also collected data about perceived identity enhancements through AI. We used these data for robustness analysis to exclude the possibility that our results regarding the effects of identity threat are biased by hidden positive attitudes. The identity enhancement items mirrored the wording of the identity threat items to capture any positive expectations toward the change induced by AI systems (Multimedia Appendix 3). The analysis shows that identity enhancement reduced resistance to AI (P=.02) and was not related to self-threat (P=.52). After including identity enhancement as a control variable, the above-described effects of perceived identity threat remained qualitatively unchanged. This indicates that perceived professional identity threats have a significant effect on resistance to AI over and beyond any effects of perceived identity enhancement through AI.

Finally, we conducted a content analysis of the qualitative statements from novice physicians to validate that our measured dimensions capture the experienced changes through AI systems. The data set consisted of a total of 414 distinct statements by 176 participants. The content of all statements was classified as positive or negative statements about AI systems by 2 independent coders (EJ and one student assistant; Table 7).

Table 7. Content analysis of the qualitative statements.

Dimensions of identity threat and categories identified in the data	Example statements from participants (survey ID)
Threats to professional recognition	
AI[a] replaces important knowledge tasks leading to a loss in professional status	• "Physicians appear to be less competent as patients believe that a small machine can also find solutions and solve their problem." (#98) • "Physicians will be replaced by computers and will only fulfill an assistant job." (#894)
Decreasing importance of professional knowledge and expertise	• "Physicians are tempted to study less as they know that the system will have all data accessible." (#492)
Threats to professional capabilities	
Loss of professional autonomy in decision-making	• "Physicians might stop to think for themselves and do not reflect on the results presented by the system which is a severe source of mistakes" (#98) • "Less legal protection for the physician if he/she acts against the AI's recommendation due to experience" (#954)
Pressuring the role of care provider	• "A physician will be more someone who operates a machine than being a doctor taking care of his/her patient's individual needs." (#2426) • "less patient oriented care when using technology" (#850)
Loss of influence	• "The management will then require even faster decisions which results in increased time pressure." (#719)
Perceived enhancements	
AI supports decision-making by increasing decision certainty	• "AI is a relief for the physician and helps to gain security in diagnosing by providing a second opinion which either encourages the physician to reflect his diagnosis a second time or strengthens the certainty of having found the right diagnosis." (#47) • "work more independently and provides them more security in their work." (#3159)
AI supports decision-making by helping to stay up-to-date	• "it is no longer necessary to know the smallest detail of every single disease." (#743) • "AI stays abreast of the fast changes and developments in medical science." (#250)
AI increases workflow efficiency	• "AI saves time which can then be invested in the treatment of more patients or more personalized care." (#94)
AI leads to better patient care	• "AI increases security in diagnosing and might lead to better results. This again can increase the patient's trust towards the physician." (#2277)

[a]AI: artificial intelligence.

One-third of the negative statements (34/105; 32.4%) described threats to professional recognition. The implementation of AI systems was perceived as leading to a loss of status and prestige for the occupational group of physicians and made participants fear that physicians might become redundant and reduced to a mere voice of the AI system. Moreover, participants feared that expert knowledge would become less important as AI systems incorporated more up-to-date knowledge than ever possible for a human being. The statements also contained multiple threats to professional capabilities. As such, participants feared that physicians might lose their autonomy in decision-making as they might trust the AI system more than appropriate, whereas the system would perform tasks autonomously. In addition, participants perceived that it would become more difficult to be a care provider with AI systems in place, as these systems would increase the distance between physicians and patients. The participants also feared that liability issues would arise if they disagreed with AI decisions. Conversely, 3 categories of positive statements emerged from the content analysis. AI systems were perceived as supporting decision-making through reduced uncertainty and complexity in diagnostics. Moreover, they were seen as facilitators of access to knowledge, supporting especially novice physicians, by providing access to the newest guidelines and empirical findings.

Discussion

Principal Findings

Our work contributes to the knowledge on the impact of AI and the future of work in health care [53-55]. It shows that professional identity threats from AI systems are indeed a serious concern for novice and experienced physicians and contribute to resistance to AI. AI systems threaten both professional recognition and professional capabilities of medical professionals. Threats to professional capabilities directly contribute to resistance to AI, whereas the effect of threats to professional recognition is mediated through self-threat. Professional experience and perceived temporal distance of AI systems influence the relationship between perceived identity threats and resistance attitudes. Medical novices experience

stronger identity threats than medical professionals. In addition, if AI systems are perceived as more relevant in the near future, threats to professional capabilities are more profound. If, however, AI systems are perceived as relevant in the distant future, threats to personal recognition gain in importance.

Our findings have implications for the understanding of how the medical professional identity changes with increasingly powerful AI systems and how AI systems are integrated into medical decisions. First, experienced identity threats influence how physicians adapt their professional identity to the upcoming change. For instance, study participants who indicated that "the role of the physician will be more passive, since decisions will be automated" might be less likely to choose specialties such as radiology. This can lead to fewer physicians who actively work with AI systems and develop the technological capabilities to evaluate those systems. Furthermore, several participants declared that they planned to focus on soft skills instead of analytical decision-making skills, which would rather be performed by an AI. Thus, instead of using AI systems as a second opinion and engaging in elaborate decision-making, physicians might end up delegating important tasks to AI systems without considering them in detail.

Second, threats to the professional identity cause identity protection responses [20] that directly impact technology use. In health care, physicians are pivotal for developing the ground truth for learning algorithms and for identifying relevant explanations and predictive values. Furthermore, physicians can make better diagnosis decisions with the support of trained algorithms and use them as a second opinion [56]. However, if they feel threatened in their identity, physicians are less likely to engage in the active development and adaptation of AI systems and resist their implementation. Moreover, identity protection responses can lead to incorrect medical decisions with AI systems if physicians reject AI advice as soon as it contradicts their opinion and is perceived as threatening [51]. In particular, threats to professional capabilities play a focal role in developing negative attitudes toward AI systems and should, thus, be addressed through specific medical training in interacting with AI systems.

Limitations and Future Research

This study has several limitations that can serve as a springboard for future research. First, by using the survey method, we were not able to capture how the identity develops through a longer period of time and whether medical students who perceived stronger threats to their future from AI would switch to a *nonthreatened* profession that requires more subjective interpersonal skills. In addition, it would be interesting to see how the professional identity is affected in clinical practice through a more intensive interaction with AI systems. Furthermore, our study provides first insights into potential differences in experienced identity threats across medical specialties. Specialties such as radiology or pathology were scarce in our sample, although those specialties often use AI in medical practice. Consequently, a follow-up study that looks at differences across specialties in more detail might provide interesting insights. In addition, our sample consisted of respondents who reported a relatively low degree of familiarity with AI systems. This reflects the current situation in medical education, in which medical novices are not trained in the use of AI systems. However, whether a sample with more familiarity would experience lower degrees of threat from AI systems needs to be further researched. Second, as noted in the literature [28,33,57] and underlined by the qualitative survey responses, there are also positive appraisals of AI systems that can enhance, rather than threaten, individuals' identity. Given that there are both strong positive and negative perceptions of the impact of AI systems on the professional identity, future research should consider the impact of ambivalence [58] on professional identity formation and restructuring. Third, we presented an AI system with a 90% accuracy rate. However, in clinical practice, the accuracy rate is highly dependent on the context, that is, the complexity of patient cases, and can be heavily disputed by medical professionals. Furthermore, with lower perceived or actual accuracy, physicians might develop more negative attitudes toward the AI system. Finally, as our study examined only 2 dependent variables, it is important to investigate how professional identity threat from AI systems impacts other variables, such as anxiety and long-term behaviors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed overview of prior research on identity threats.
[DOCX File , 31 KB - formative_v6i3e28750_app1.docx]

Multimedia Appendix 2
Overview of item development process.
[DOCX File , 38 KB - formative_v6i3e28750_app2.docx]

Multimedia Appendix 3
Additional analysis details, including confirmatory factor analysis, common method bias analysis, regression analysis details for models 2 and 3, robustness analysis with identity enhancement, and details on identity enhancement items.
[DOCX File , 41 KB - formative_v6i3e28750_app3.docx]

References

1. Russell S, Norvig P. Artificial Intelligence: A Modern Approach. Upper Saddle River: New Jersey: Pearson Education Limited; 2010.

2. Mayo RC, Leung J. Artificial intelligence and deep learning - Radiology's next frontier? Clin Imaging 2018 May;49:87-88. [doi: 10.1016/j.clinimag.2017.11.007] [Medline: 29161580]

3. Shen J, Zhang CJ, Jiang B, Chen J, Song J, Liu Z, et al. Artificial intelligence versus clinicians in disease diagnosis: systematic review. JMIR Med Inform 2019 Aug 16;7(3):e10010 [FREE Full text] [doi: 10.2196/10010] [Medline: 31420959]

4. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJ. Artificial intelligence in radiology. Nat Rev Cancer 2018 Aug 17;18(8):500-510 [FREE Full text] [doi: 10.1038/s41568-018-0016-5] [Medline: 29777175]

5. Jha S, Topol EJ. Adapting to artificial intelligence: radiologists and pathologists as information specialists. JAMA 2016 Dec 13;316(22):2353-2354. [doi: 10.1001/jama.2016.17438] [Medline: 27898975]

6. Baird A, Maruping LM. The next generation of research on IS use: a theoretical framework of delegation to and from agentic IS artifacts. MIS Q 2021;45(1):315-341. [doi: 10.25300/misq/2021/15882]

7. Willcocks L. Robo-Apocalypse cancelled? Reframing the automation and future of work debate. J Inf Technol 2020 Jun 10;35(4):286-302. [doi: 10.1177/0268396220925830]

8. Tang A, Tam R, Cadrin-Chênevert A, Guest W, Chong J, Barfett J, Canadian Association of Radiologists (CAR) Artificial Intelligence Working Group. Canadian association of radiologists white paper on artificial intelligence in radiology. Can Assoc Radiol J 2018 May 01;69(2):120-135 [FREE Full text] [doi: 10.1016/j.carj.2018.02.002] [Medline: 29655580]

9. Bhattacherjee A, Hikmet N. Physicians' resistance toward healthcare information technology: a theoretical model and empirical test. Eur J Inf Syst 2017 Dec 19;16(6):725-737. [doi: 10.1057/palgrave.ejis.3000717]

10. Lapointe L, Rivard S. A multilevel model of resistance to information technology implementation. MIS Quarterly 2005;29(3):461-491. [doi: 10.2307/25148692]

11. Abdullah R, Fakieh B. Health care employees' perceptions of the use of artificial intelligence applications: survey study. J Med Internet Res 2020 May 14;22(5):e17620 [FREE Full text] [doi: 10.2196/17620] [Medline: 32406857]

12. Oh S, Kim JH, Choi S, Lee HJ, Hong J, Kwon SH. Physician confidence in artificial intelligence: an online mobile survey. J Med Internet Res 2019 Mar 25;21(3):e12422 [FREE Full text] [doi: 10.2196/12422] [Medline: 30907742]

13. Blease C, Locher C, Leon-Carlyle M, Doraiswamy M. Artificial intelligence and the future of psychiatry: qualitative findings from a global physician survey. Digit Health 2020;6:2055207620968355 [FREE Full text] [doi: 10.1177/2055207620968355] [Medline: 33194219]

14. Chreim S, Williams BE, Hinings CR. Interlevel influences on the reconstruction of professional role identity. Acad Manag J 2007 Dec 1;50(6):1515-1539. [doi: 10.5465/amj.2007.28226248]

15. Stets JE, Burke PJ. Identity theory and social identity theory. Soc Psychol Q 2000 Sep;63(3):224-237. [doi: 10.2307/2695870]

16. Tajfel H. Social psychology of intergroup relations. Annual Rev Psychol 1982 Jan;33(1):1-39. [doi: 10.1146/annurev.ps.33.020182.000245]

17. Pratt MG, Rockmann KW, Kaufmann JB. Constructing professional identity: the role of work and identity learning cycles in the customization of identity among medical residents. Acad Manag J 2006 Apr;49(2):235-262. [doi: 10.5465/amj.2006.20786060]

18. Hoff TJ. The physician as worker: what it means and why now? Health Care Manage Rev 2001;26(4):53-70. [doi: 10.1097/00004010-200110000-00006] [Medline: 11721310]

19. Reay T, Goodrick E, Waldorff SB, Casebeer A. Getting leopards to change their spots: co-creating a new professional role identity. AMJ 2017 Jun;60(3):1043-1070. [doi: 10.5465/amj.2014.0802]

20. Petriglieri J. Under threat: responses to and the consequences of threats to individuals' identities. AMR 2011 Oct;36(4):641-662. [doi: 10.5465/amr.2009.0087]

21. Murtagh N, Gatersleben B, Uzzell D. Self-identity threat and resistance to change: evidence from regular travel behaviour. J Environ Psychol 2012 Dec;32(4):318-326. [doi: 10.1016/j.jenvp.2012.05.008]

22. Seibt B, Förster J. Stereotype threat and performance: how self-stereotypes influence processing by inducing regulatory foci. J Pers Soc Psychol 2004 Jul;87(1):38-56. [doi: 10.1037/0022-3514.87.1.38] [Medline: 15250791]

23. Campbell WK, Sedikides C. Self-threat magnifies the self-serving bias: a meta-analytic integration. Rev General Psychol 1999 Mar 01;3(1):23-43. [doi: 10.1037/1089-2680.3.1.23]

24. Sassenberg K, Sassenrath C, Fetterman AK. Threat ≠ prevention, challenge ≠ promotion: the impact of threat, challenge and regulatory focus on attention to negative stimuli. Cogn Emot 2015;29(1):188-195. [doi: 10.1080/02699931.2014.898612] [Medline: 24650166]

25. Stein M, Newell S, Wagner EL, Galliers RD. Coping with information technology: mixed emotions, vacillation, and nonconforming use patterns. MIS Q 2015 Feb 2;39(2):367-392. [doi: 10.25300/misq/2015/39.2.05]

26. Craig K, Thatcher J, Grover V. The IT identity threat: a conceptual definition and operational measure. J Manag Inf Syst 2019 Mar 31;36(1):259-288. [doi: 10.1080/07421222.2018.1550561]

27. Kyratsis Y, Atun R, Phillips N, Tracey P, George G. Health systems in transition: professional identity work in the context of shifting institutional logics. Acad Manag J 2017 Apr;60(2):610-641. [doi: 10.5465/amj.2013.0684]

28.   Mishra AN, Anderson C, Angst CM, Agarwal R. Electronic health records assimilation and physician identity evolution: an identity theory perspective. Inf Syst Res 2012 Sep;23(3-part-1):738-760. [doi: 10.1287/isre.1110.0407]

29.   Walter Z, Lopez MS. Physician acceptance of information technologies: role of perceived threat to professional autonomy. Decision Support Syst 2008 Dec;46(1):206-215. [doi: 10.1016/j.dss.2008.06.004]

30.   Esmaeilzadeh P, Sambasivan M, Kumar N, Nezakati H. Adoption of clinical decision support systems in a developing country: antecedents and outcomes of physician's threat to perceived professional autonomy. Int J Med Inform 2015;84(8):548-560. [doi: 10.1016/j.ijmedinf.2015.03.007] [Medline: 25920928]

31.   Doolin B. Power and resistance in the implementation of a medical management information system. Inform Syst J 2004 Oct;14(4):343-362. [doi: 10.1111/j.1365-2575.2004.00176.x]

32.   Nach H. Identity under challenge. Manag Res Rev 2015;38(7):703-725. [doi: 10.1108/MRR-02-2014-0031]

33.   Jensen TB, Aanestad M. Hospitality and hostility in hospitals: a case study of an EPR adoption among surgeons. Eur J Inf Syst 2017;16(6):672-680. [doi: 10.1057/palgrave.ejis.3000713]

34.   Korica M, Molloy E. Making sense of professional identities: stories of medical professionals and new technologies. Human Relations 2010 Sep 23;63(12):1879-1901. [doi: 10.1177/0018726710367441]

35.   Zimmermann A, Ravishankar M. Collaborative IT offshoring relationships and professional role identities: reflections from a field study. J Vocational Behav 2011 Jun;78(3):351-360. [doi: 10.1016/j.jvb.2011.03.016]

36.   Luguri JB, Napier JL. Of two minds: the interactive effect of construal level and identity on political polarization. J Experimental Soc Psychol 2013 Nov;49(6):972-977. [doi: 10.1016/j.jesp.2013.06.002]

37.   MacKenzie SB, Podsakoff PM, Podsakoff NP. Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques. MIS Q 2011;35(2):293-334. [doi: 10.2307/23044045]

38.   Willis GB, Royston P, Bercini D. The use of verbal report methods in the development and testing of survey questionnaires. Appl Cognit Psychol 1991 May;5(3):251-267. [doi: 10.1002/acp.2350050307]

39.   Fisher RJ. Social desirability bias and the validity of indirect questioning. J Consum Res 1993 Sep;20(2):303-315. [doi: 10.1086/209351]

40.   Bick M, Kummer T, Ryschka S. Determining anxieties in relation to ambient intelligence—explorative findings from hospital settings. Inf Syst Manag 2015 Jan 08;32(1):60-71. [doi: 10.1080/10580530.2015.983021]

41.   Moore GC, Benbasat I. Development of an instrument to measure the perceptions of adopting an information technology innovation. Inf Syst Res 1991 Sep;2(3):192-222. [doi: 10.1287/isre.2.3.192]

42.   Jarvis C, MacKenzie S, Podsakoff P. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. J Consum Res 2003 Sep;30(2):199-218. [doi: 10.1086/376806]

43.   Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Market Res 2018 Nov 28;18(1):39-50. [doi: 10.1177/002224378101800104]

44.   van der Heijden H. User acceptance of hedonic information systems. MIS Quarterly 2004;28(4):695-704. [doi: 10.2307/25148660]

45.   Iacobucci D. Structural equations modeling: fit Indices, sample size, and advanced topics. J Consumer Psychol 2010 Jan;20(1):90-98. [doi: 10.1016/j.jcps.2009.09.003]

46.   Straub D, Gefen D. Validation Guidelines for IS Positivist Research. CAIS 2004;13:380-427. [doi: 10.17705/1CAIS.01324]

47.   Podsakoff PM, MacKenzie SB, Lee J, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]

48.   Ho SY, Bodoff D. The effects of web personalization on user attitude and behavior: an integration of the elaboration likelihood model and consumer search theory. MIS Q 2014 Feb 2;38(2):497-520. [doi: 10.25300/misq/2014/38.2.08]

49.   Liang H, Saraf N, Hu Q, Xue Y. Assimilation of enterprise systems: the effect of institutional pressures and the mediating role of top management. MIS Quarterly 2007;31(1):59. [doi: 10.2307/25148781]

50.   Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]

51.   Jussupow E, Spohrer K, Heinzl A, Gawlitza J. Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence. Inf Syst Res 2021 Sep;32(3):713-735. [doi: 10.1287/isre.2020.0980]

52.   Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York, USA: Guilford; 2013.

53.   Verghese A, Shah NH, Harrington RA. What this computer needs is a physician: humanism and artificial intelligence. JAMA 2018 Jan 02;319(1):19-20. [doi: 10.1001/jama.2017.19198] [Medline: 29261830]

54.   Fazal MI, Patel ME, Tye J, Gupta Y. The past, present and future role of artificial intelligence in imaging. Eur J Radiol 2018;105:246-250. [doi: 10.1016/j.ejrad.2018.06.020] [Medline: 30017288]

55.   Fichman RG, Kohli R, Krishnan R. The role of information systems in healthcare: current research and future trends. Inf Syst Res 2011;22(3):419-428. [doi: 10.1287/isre.1110.0382]

56.   Cheng J, Ni D, Chou Y, Qin J, Tiu C, Chang Y, et al. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in CT scans. Sci Rep 2016 Apr 15;6(1):24454 [FREE Full text] [doi: 10.1038/srep24454] [Medline: 27079888]

57.   Stein M, Galliers RD, Markus ML. Towards an understanding of identity and technology in the workplace. J Inf Technol 2013 Sep 01;28(3):167-182. [doi: 10.1057/jit.2012.32]
58.   Piderit SK. Rethinking resistance and recognizing ambivalence: a multidimensional view of attitudes toward an organizational change. Acad Manag Res 2000 Oct;25(4):783-794. [doi: 10.5465/amr.2000.3707722]

Abbreviations

AI: artificial intelligence

XSL•FO
RenderX

<u>Original Paper</u>

Designing a Cancer Prevention Collaborative Goal-Setting Mobile App for Non-Hispanic Black Primary Care Patients: An Iterative, Qualitative Patient-Led Process

Daniel Resnick[1], MD; Matthew D Kearney[2], MPH, PhD; Jazmine M Smith[3,4], MSc; Allison Bautista[3,4], BSN; Liz Jones[5], BFA; Marilyn M Schapira[4,6,7], MPH, MD; Jaya Aysola[3,4,6,8], MD, DTMH, MPH

[1]Department of Medicine, Emory University School of Medicine, Atalnta, GA, United States

[2]Department of Family Medicine and Community Health, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[3]Penn Medicine Center for Health Equity Advancement, Office of the Chief Medical Officer, University of Pennsylvania Health System and Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[4]Division of General Internal Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[5]Transmogrify, Conshohocken, PA, United States

[6]Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA, United States

[7]The Center for Health Equity Research and Promotion, Michael J Crescenz Veterans Administration Medical Center, Philadelphia, PA, United States

[8]Office of Inclusion and Diversity, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Jaya Aysola, MD, DTMH, MPH
Office of Inclusion and Diversity
Perelman School of Medicine
University of Pennsylvania
423 Guardian Drive
Philadelphia, PA, 19104
United States
Phone: 1 215 573 9475
Email: jaysola@upenn.edu

Abstract

Background: There remains a need to engage at-risk primary care populations in cancer prevention behaviors, yet primary care physicians often lack the time or resources to discuss these behaviors with their patients.

Objective: The objective of this study is to evaluate the content, usability, and acceptability of a mobile app that leverages insights from goal-setting and social network literature to facilitate cancer prevention goal setting, tracking, and sharing between non-Hispanic Black primary care patients and their social ties.

Methods: We recruited eligible non-Hispanic Black primary care patients (aged ≥18 years) from 2 practice sites in West Philadelphia, using nonprobabilistic purposive sampling. We conducted semistructured interviews with 5 to 7 participants over 3 weeks to solicit feedback on paper mock-ups of the app, iteratively adapting these mock-ups after each set of interviews. Thereafter, and informed by initial feedback, we created an electronic beta version of the app and sought acceptability and usability feedback from a different set of participants. Then, we conducted content analysis of all user responses to search for unifying themes on acceptability and usability of both the initial mock-ups and beta version of the app. We further assessed app usability using questions derived from the System Usability Scale.

Results: A total of 33 non-Hispanic Black primary care patients participated in this study. The mean age was 49 (SD 13) years, and 26 (79%) out of 33 participants identified as female. Semistructured interviews revealed three primary generalizable insights from our target population: the framing of each goal and its relevance to cancer impacted the likelihood that the goal would be chosen, participants thought that sharing health goals with others facilitates health behaviors, and most participants found it motivating to see other users' goal progress, while still collaborating with these users on their health goals. An overarching insight that permeated across each theme was the participants' desire to customize and personalize the app. Usability testing revealed that 100% (33/33) of participants found the app easy to use, and 76% (25/33) of participants reported that they would like to use this app frequently.

XSL•FO
RenderX

Conclusions: Cancer prevention in the modern era must include options that are accessible to all, but this does not mean that all options must be universal. This study's iterative process led to the development of a cancer prevention mobile app that non-Hispanic Black primary care patients deemed usable and acceptable and yielded noteworthy insights about what intended end users value in setting and accomplishing health goals.

Introduction

Background

Increasing the adoption of health behaviors at a population level is essential if we are to significantly decrease the burden of preventable cancer and improve public health. In the United States, more than 600,000 people die of cancer each year [1]. Approximately 30% of these deaths are linked to poor diet, physical inactivity, and carrying too much weight, with another 30% due to tobacco use, comprising nearly two-thirds of US cancer deaths [2-5]. Furthermore, cancer disproportionately impacts non-Hispanic Black populations largely due to inequities stemming from structural racism [6-9]. Evidence suggests that current primary care services do not effectively engage all patients in cancer prevention [10-16]. Therefore, there remains a need for other potential interventions to address this gap, especially among those most at risk.

One strategy to increase cancer prevention health behaviors is goal setting [17,18]. Collaborative goal setting, a process whereby the provider and patient agree upon a health-related specific SMART (specific, measurable, achievable, realistic, and time-bound) goal and action plan [19], has been shown to modify behaviors by directing intention and building self-efficacy [20-22]. However, in primary care, we lack an approach to implement a strategy for collaborative goal setting. A second approach to increase cancer prevention health behaviors is to disseminate health behaviors and knowledge through social networks, which are known to influence behaviors related to cancer risk, such as obesity and smoking [23-26]. Experimental studies suggest that reinforcement from multiple social ties (ie, through a network) increases health behavior adoption more than social reinforcements from single ties [27]. Prior work also suggests that cancer prevention strategies involving some form of social support are more effective in changing behaviors in BIPOC (Black, Indigenous, and People of Color) populations as compared with non-Hispanic White people [28,29]. There is evidence that BIPOC populations have denser social networks, with more reliable and frequent activation of informal social support [12,14].

The objective of this study is to develop and evaluate the content, usability, and acceptability of an electronic decision support tool—ie, a mobile app—that leverages these insights to facilitate cancer prevention goal setting, tracking, and sharing between primary care patients and their social ties. We conducted a series of semistructured interviews to determine the optimal content and app features before piloting the prototype with our priority population: non-Hispanic Black primary care patients. Although non-Hispanic Black populations use health technology at greater rates than their White counterparts [30-32], they remain underrepresented in studies about health technology and health behaviors [33,34]. There is evidence that end user experiences may vary by background and culture with the need for culturally sensitive and effective design [35-37]. In addition, there is a call for evaluating public health interventions with messaging grounded in the understanding of the populations served and without White bias [38,39].

Objective

Given that this app will center on facilitating cancer prevention behaviors in populations most at risk, we aimed to ensure the app is culturally attuned to and meets the needs of its targeted end user. We also aimed to test features, such as leveraging social ties, given the evidence that such features may work better among non-Hispanic Black populations [28,29]. Therefore, this study aims to evaluate the content, usability, and acceptability of a cancer prevention app designed with direct input from and specifically for a non-Hispanic Black primary care patient population.

Methods

Study Overview

We conducted a multistage, mixed methods study to develop and evaluate the content, usability, and acceptability of a mobile app that facilitates cancer prevention goal setting, goal tracking, and goal sharing. Consistent with mHealth app development best practices [40], this study comprised 2 stages. First, we obtained feedback from potential end users regarding the paper prototypes of the app. Then, we solicited feedback on a beta version of the app, informed by initial feedback, from a new set of participants. Participants completed questionnaires containing both open- and closed-ended questions, which were subsequently analyzed to refine the prototype. The study team guided the design, features, and content of both the paper prototype and the electronic beta version of the mobile app in collaboration with Transmogrify (Conshohocken, PA), a firm that helps create, build, and grow digital products. Figure 1 shows a visual representation of the study design and stages.

Figure 1. Visual representation of the app design process and the iterative changes made during app development.
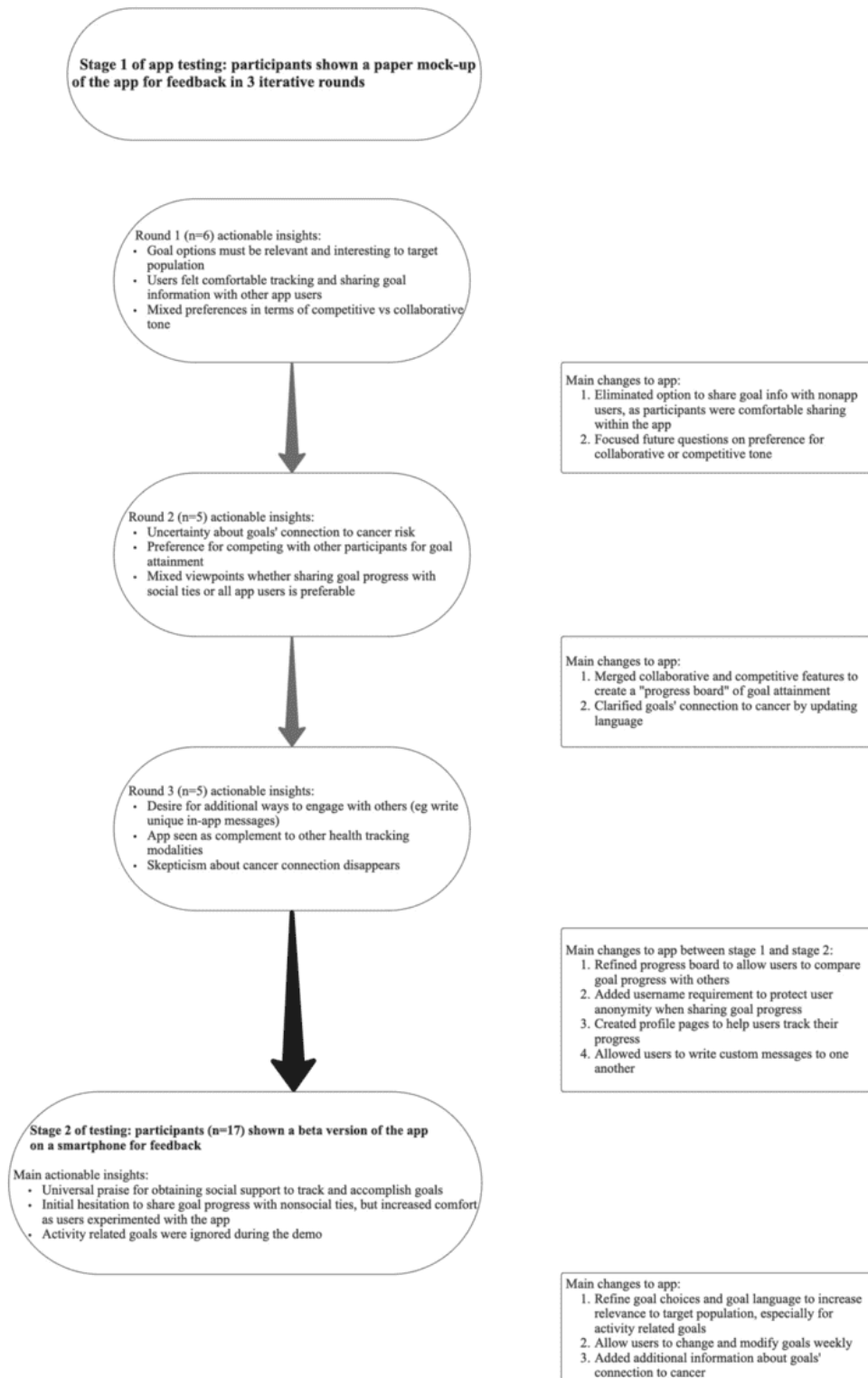
Setting and Participants

Participants and Eligibility Criteria

Participants had to (1) self-identify as non-Hispanic Black, (2) be aged ≥18 years, (3) speak English, (4) be seen at one of our study sites once in the past 3 years (if the patient has a designated primary care provider [PCP]) or twice in the past 2 years (if no assigned PCP), (5) be able to provide informed

consent, and (6) not have participated in another stage of testing for the intervention.

Setting and Recruitment Process

We recruited participants for both stages at 2 internal medicine primary care clinics in West Philadelphia, which serve a racially/ethnically diverse patient population, using a nonprobabilistic purposive sampling technique. We first generated a list of patients that met the first 4 eligibility criteria

above with upcoming appointments. Then, between April and May 2019 (stage 1) and May and June 2019 (stage 2), study team members invited potential participants from the clinics' waiting rooms that were on the pregenerated list to screen for the study. Interested participants reported to a private room after their appointments, where the study team confirmed the participants' eligibility, including their self-reported race/ethnicity, informed participants of the study's aims, and obtained formal consent for participation.

In stage 1, we aimed to recruit 5 to 7 participants per week over 3 weeks to rapidly modify the prototype based on participant feedback. Early usability testing research demonstrates that optimal feedback is derived from multiple rounds of testing with potential end users that inform refinements in between rounds [41-43], rather than one larger study that examines only one version of an app. In stage 2, we targeted a sample size of 15 to 20 participants to achieve thematic saturation of feedback and generate quantitative usability data of the beta version of the app [42,44]. Participants were incentivized at US $30 to complete the interviews.

Ethics Approval

The University of Pennsylvania Institutional Review Board approved the protocol for this study (828151).

Mobile App Prototype

The main objectives of the mobile app prototype were to (1) communicate the value of collaborative goal setting for cancer prevention, (2) provide a selection of concrete SMART goals [45] related to cancer prevention behaviors informed by evidence-based guidelines and recommendations from the American Cancer Society (ACS) [46], and (3) serve as a patient-held prompt to facilitate cancer prevention collaborative goal-setting discussions with PCPs and encourage easy sharing of information with social ties. The earliest version of the prototype was based on guidelines for the adoption of cancer prevention behaviors [29]. We ensured it tested at a Flesch Reading Ease score of 86.2% (ie, is understood by 11- to 13-year-olds) and a Flesch-Kincaid grade level score of 4.3 (ie, is at a fourth-grade reading level) [47]. Through quantitative and qualitative assessments, we aimed to solicit end user perspectives and preferences on the apps' (1) content and format, (2) delivery and use during primary care visits, and (3) use to share information with social ties (please refer to Multimedia Appendices 1 and 2 for iterative versions of our prototype at different stages of our study).

Data Collection, Measurements, and Analysis

Qualitative and Quantitative Data Collection

Stage 1

Over a 3-week period, the study team conducted semistructured interviews with 5 to 7 participants each week to solicit feedback on paper mock-ups of the app. Semistructured interviews averaged 30 minutes and included both closed and open-ended questions about the app's content, features, delivery methods, and potential future use. The study team synthesized and discussed the interview feedback weekly, iteratively refining the prototype before the subsequent set of participant interviews

until the study team felt as though it could proceed to the next stage of testing. Please refer to Multimedia Appendix 1 for examples of our prototype and the corresponding interview guides.

Stage 2

Before stage 2, the development team transformed the latest paper prototype into an electronic beta version of the app. In stage 2, the study team walked individual participants (n=17) through the beta version of the app on a smartphone. During this 30-minute walk-through, team members asked each participant approximately 33 close-ended and approximately 20 open-ended questions about the usability and acceptability of the app and its features. Multimedia Appendix 2 illustrates an example of our prototype and interview guide.

For both stages, we recorded all open-and close-ended responses verbatim into REDCap (Research Electronic Data Capture; REDCap Consortium) [48].

Participant Characteristics

In stage 1, we assessed participants (n=16), age, sex, technology use, and health habits. Survey questions evaluated participants' typical use of their mobile devices, comfortability with sharing health information with social ties on the internet and offline, and current goal setting and health tracking behaviors. In stage 2 (n=17), we collected the participants' age and sex.

Qualitative Measures

We asked certain open-ended questions in both study stages to inquire about the participants' overall impressions of the acceptability and usability of the app and its main features. We modified the interview guide iteratively each week in stage 1 to incorporate questions focused on specific changes made to the app based on the prior week's feedback. Examples of our early prototypes and corresponding interview scripts are provided in Multimedia Appendix 1. The stage 1 prototype went through 10 refinements to inform stage 2. For stage 2, we developed this prototype within InVision [49], a digital product design platform that allows end users to interact with the prototype as if it is an app. Examples of this prototype and sample interview guides are provided in Multimedia Appendix 2.

Quantitative Measures

In addition to the open-ended questions, we asked participants in stage 2 (n=17) close-ended survey questions about the usability of certain app features and the app overall. These questions were adapted from the System Usability Scale [50], an instrument commonly used to evaluate the usability of different technology products.

Analysis

We first assessed the participants' characteristics by tabulating the distributions or frequencies of the questions detailed above. We also calculated the distribution of Likert scale responses, ranging from strongly disagree to strongly agree, for the modified *System Usability Scale* questions asked in stage 2 of app testing.

Iterative Qualitative Analysis

Study team members (JMS, AB, LJ, and JA) met after each round of interviews to analyze feedback for key themes to inform refinements to the app content and features. This form of analysis allowed for the rapid implementation of the participants' feedback and strengthened the development of the app [42].

Qualitative Content Analysis

Two team members (DR and MDK) conducted a qualitative content analysis [51] of all responses to search for unifying themes across stages of testing, by reading through the responses and creating the initial codebook. We randomly selected one interview from each round of testing (4 total) to refine the codebook and achieve consensus on code definition, inclusion criteria, and exclusion criteria. JMS then coded the remaining interviews using a constant-comparison technique. Throughout the coding process, a total of 8 out of 33 interviews (24% of the total sample) were jointly coded by DR, MDK, and JMS to assess interrater reliability. Facilitated by NVivo software (version 12; QSR International), we calculated the percent agreement [52], a measure of coding consensus, and determined that there was satisfactory interrater reliability (median 75% agreement; mean 69.4%, SD 23.5%). The study team then reviewed all coded responses and extracted key themes from across responses. This allowed for the simultaneous analysis of interviews collected throughout all stages of app development.

Results

Participant Characteristics

Of the 33 non-Hispanic Black primary care patients participating in the study, the mean age was 49 (SD 13) years, and 26 (79%) identified as female. Of the initial stage 1 (n=16) participants, 14 (88%) reported using a smartphone multiple times a day, 7 (44%) specifically used an app or digital fitness tracker to track their health, 13 (81%) reported tracking their health either digitally or manually, 15 (94%) reported sharing "some" or "a lot" of health information with close friends and family, and 13 (81%) said they have relied on friends to accomplish health goals. Only 25% (4/16) of the participants reported that they were comfortable discussing health matters on the internet.

Qualitative Analysis

Table 1 summarizes the 3 dominant themes and associated subthemes that emerged from the qualitative content analysis of both stages. Below, we expand on these qualitative themes and add additional insights and changes made during the iterative analysis.

Table 1. Content themes and representative quotes.

Theme and subtheme	Representative quote
Messaging matters	
SMART[a] goals resonate	"It's something that I'm already working on, so actually a lot of [the] options were pretty good, so I wanted to pick more than one of them" [stage 2]
Achieving buy-in for cancer prevention messaging	"...[the app is] to the point. It tells me exactly what we're working on and gives me some things right on hand to reduce my chances of getting cancer." [stage 1, round 2]
Specifying goals for the target population	"I was already interested in cutting down red meat, but I wasn't sure if I was ready to do it yet. So, it was cool to see that as an option." [stage 2]
To share or not to share	
Working with others facilitates goal accomplishment	"I like the idea of sharing with friends and family and seeing other people sharing their progress. Overall, I think it's pretty good. It helps you keep on track." [stage 2]
Preferences for sharing goals with loved ones only versus all app users	"...I like [the app]. I would only pick [to share with] my friends. Since I'm trying to quit smoking, I wouldn't open it to everyone." [stage 2];
	"Family sometimes are critical. You could get more compassion from someone you don't know." [stage 1, round 2]
Competition versus collaboration	
Deriving motivation from competition	"Ah yeah, some people like to do things out of competition." Prompt: Would it be motivating for you personally? "Yes. I don't like to lose." [stage 1, round 2]
Success through collaboration	"I think it's a good idea to be able to communicate with [other users] the things that they are doing and the things that I'm doing to make better choices to reduce our risk of contracting cancer." [stage 2]
The progress board: a Goldilocks solution	"I really like the progress [board]. I like that you can click on a person and send them encouragement, or even your own personal message. I think I would use this app." [stage 2]

[a]SMART: specific, measurable, achievable, realistic, and time-bound.

Theme 1: Messaging Matters

SMART Goal Framing Resonates

A key objective of this app is to encourage the setting of appropriate and motivating health goals. Therefore, we sought feedback on how to present the SMART goals in a manner that is relevant to the target population, which highlights the important connection between these goals and cancer prevention. Overall, we found that messaging mattered to the participants. Participants appreciated the goals' specificity and commented that these goals seem relevant to their efforts to become healthier. One participant stated as follows:

> [The goals are] good...they're all something that I can work on. I like the 'Make small changes for a big impact.' That makes a lot of sense. [stage 1, round 3]

### Achieving Buy-in for Cancer Prevention Messaging

A few participants noted confusion and skepticism about the relationship between lifestyle behaviors and cancer prevention. One participant said as follows:

> only the smoking makes people think of cancer. The other ones seem more like basic health as opposed to associating it with cancer.

She then asked for "more specific information about how [these goals] relate to cancer." [stage 1, round 2]. This initial feedback led to the following modifications: (1) we changed the language in the app to emphasize that small lifestyle modifications, such as those advocated by the app, can lead to a direct impact on cancer risk and (2) we added direct links to ACS webpages on lifestyle behaviors and cancer, to further emphasize the importance of these goals. Feedback on these modifications was positive in stage 2, with 15 (88%) participants out of 17 indicating that they agreed or strongly agreed that the connection to cancer in the app was both clear and useful. (See Figure 2 for additional quantitative results from stage 2).

Figure 2. Usability feedback for key app features.
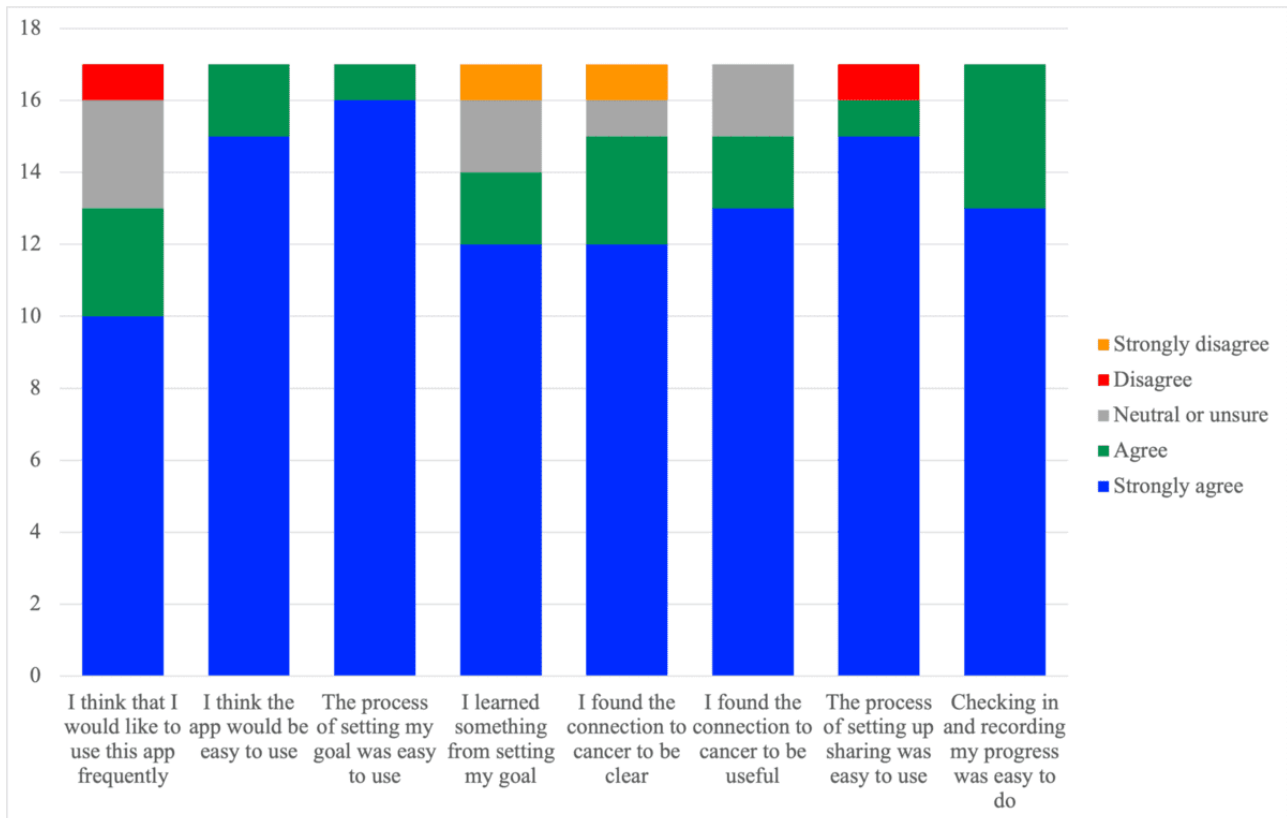
Specifying Goals for the Target Population

In another example of the importance of SMART goal framing, we made significant changes to the app's "Get Active" goals when we realized that none of the participants chose those goals when using the app in stage 2. Specifically, we centered each "Get Active" goal around a type of exercise (eg, "I will do 30 minutes of dancing"), rather than a more general goal (eg, "I will do 30 minutes of moderate intensity exercise"). As noted in a separate report, these new "Get Active" goals became more popular in future rounds of testing [53].

Desire for Customization

Participants broadly supported the goal choices offered by the app. However, nearly all participants wished to further adapt the goals themselves. Given this feedback, we added a number of customization options, from allowing users to choose the frequency of an action (eg, "I will do 30 minutes of dancing 4 times a week") to permitting users to change their goals on a weekly basis. Even with these modifications, all goals remained SMART (ie, specific, measurable, achievable, realistic, and time-bound) and connected to cancer prevention.

Theme 2: To Share or Not To Share

Working With Others Facilitates Goal Accomplishment

One key feature of this app is sharing health goals with other users and working on those goals together. Across the rounds of testing, participants universally agreed that this social component of the app was a valuable feature. One end user said as follows:

> *What I really like is the whole concept of sharing with someone else and getting them actively involved. It reminded me more of a safety plan [in the context of social work.] This is how we help you get where you need to get* [stage 1, round 2]

Participants expressed numerous benefits to working on health goals with others, from increased accountability for one's own goals to the positive consequences of helping others.

Preferences for Sharing Goals With Loved Ones Only Versus All App Users

However, there is a lack of consensus about the user with whom health goals can be shared. Many users preferred only sharing their goal progress with the users they knew before joining the app. One participant did not want to reveal that he smokes outside his social circle, while another thought she would feel "pressured" by sharing her goals with all users. Nonetheless, some participants appreciated the opportunity to work on these health goals with all app users, with a couple of participants remarking that they may receive more valuable feedback from a user they did not know rather than from close friends and family. In addition, we noticed that as participants used the app in stage 2 of testing, their willingness to share the information with all app users increased. While using the app, 47% (8/17) participants selected "share with everyone." In a postuse questionnaire, however, 82% (14/17) participants said they would select "share with everyone" in the future.

Customization Supports Personal Sharing Preferences

In response to different preferences, we modified the app to facilitate all aspects of goal sharing. Between stages 1 and 2, we added a username feature to protect anonymity when users shared goal information with unknown social ties on the app. We also permitted users to customize and change their share settings, allowing participants to choose whether to share information with only social ties or with all app users.

Theme 3: Competition Versus Collaboration

Deriving Motivation From Competition

A third key feature of the app is helping users track their health goal progress, which participants universally agreed would facilitate goal accomplishment. Many participants also commented that it was motivating to view the progress of other users on the app. In fact, a number of participants suggested that the app should create a leaderboard to foster competition among users to rise to the top:

> *I definitely like the challenging [competitive version]. It's good because it's just like a game.* [stage 1, round 2]

Success Through Collaboration

Some participants strongly rejected the idea of a competitive tone on the app, with one woman stating as follows:

> *I'm not in competition with [other users] for my health...I don't see where being in competition with someone else [is helpful]; there's certain things I'm not competitive about and my health is one of them.* [stage 1, round 3]

Instead, many participants wanted the app to facilitate collaboration on health goals among users. Participants thought the app would be a valuable space to provide and receive suggestions on how to accomplish certain goals (eg, sharing recipes for healthy meals) and encouragement for goal progress.

The Progress Board: A "Goldilocks" Solution

In synthesizing this feedback on competition and collaboration, we changed the name of the "leaderboard" to "progress board," so that users could visualize other users' health goal progress without incentivizing competition. We also added a messaging feature because users expressed a strong interest in collaborating with one another. These changes resonated with participants in stage 2 of testing, with 88% (15/17) participants stating they would check their social ties' goal progress weekly and 82% (14/17) stating they would send their friends encouraging messages weekly as well.

Discussion

Principal Findings

We conducted 2 stages of semistructured interviews with non-Hispanic Black primary care patients to develop and iteratively refine a cancer prevention goal setting mobile app. Our study yielded three primary generalizable insights from our target population: (1) the framing of each goal and its relevance to cancer impacted the likelihood that the goal would be chosen, (2) participants thought that sharing health goals with others facilitates the adoption of healthy behaviors, and (3) most participants found it motivating to see other users' goal progress, while still collaborating with these users on their health goals. An overarching insight that emerged across themes was the participants' desire to customize and personalize the app.

Our first theme highlights the importance of framing goals as relevant to cancer prevention. Some participants in our study initially struggled to understand the connection between cancer and lifestyle behaviors, remarking instead that they felt as though "genes" and "God's will" had a larger role to play. Nonetheless, we maintained the overall cancer prevention framework of the app, as health behavior research has demonstrated that framing behaviors as cancer preventing increases their adoption [54-56]. Moreover, a 2019 meta-analysis found that underserved populations in the United States are comfortable receiving cancer prevention information and interventions on the internet and through mobile devices [57]. In response to the feedback received, we clarified the language on the app to better communicate how changes in behavior can make a difference in cancer risk. We re-enforced this message with links to the ACS lay resources. Following these changes,

participants in subsequent rounds did not convey similar confusion, and in fact, many appreciated the link between health behavior and cancer prevention. This change is one example of the numerous adaptations we made to optimize the app for the intended end users.

The second and third themes indicate the values of goal sharing and collaboration, respectively. Although some participants hoped that the app would foster competition among users for completing health goals, most expressed a strong desire for the app to facilitate collaboration with others to achieve their health goals. This tension between competition and collaboration is frequently studied in behavioral economics, with mixed evidence as to which approach yields a greater impact on health behavior [58]. To respect the diverging opinions among our participants, we created a progress board that displays each user's goal success count, allowing more competitive-minded users a chance to compare their progress to others. We removed any references to "leaderboards" and added new avenues for participants to communicate with each other on the app, responding to most participants who expressed a desire to collaborate with other app users. Brewer et al [59] similarly found a "sharing board" to be a popular feature in their app promoting cardiovascular health among church-going African Americans. Participants in Brewer's study were likewise motivated by the ability to send and receive encouraging messages while keeping track of other users' progress.

Finally, we found that participants uniformly valued the ability to customize the app to meet their unique needs, and in response, we provided additional options for goal sharing and goal setting. The desire for customization is common in health app development [60,61], and we aimed to satisfy participant requests while ensuring that the app remained grounded in evidence-based techniques. For example, although we allowed participants to modify the frequency of a particular health goal (eg, dance for 30 minutes three vs four times a week), we did not allow users to write their own health goals. We made this choice to ensure that all of our goals remained consistent with ACS recommendations and an evidence-based SMART [19] set-up.

Given that this study evaluated end user experience with a mobile app exclusively among a non-Hispanic Black population, it is difficult to ascertain whether our findings are unique to this population or perhaps more broadly applicable to other populations. However, we did find a preference for collaboration over competition that aligns with prior evidence of this among non-Hispanic Black populations [59]. In addition, our findings of fatalism in early iterations of app content feedback aligned with prior studies that demonstrate this among non-Hispanic Black populations [62-64]. Future studies should be designed to compare the experiences of apps and messaging between users to determine what features may be uniquely appreciated by one racial/ethnic population as compared with another.

Strengths and Limitations

These findings must be considered in the context of several limitations and strengths. First, the study used a nonprobabilistic purposive sampling technique to recruit non-Hispanic black primary care patients from 2 primary care clinics and thus may not be generalizable. Second, we recognize that stated intentions do not always align with future behavior, and we cannot predict the effectiveness of this tool based on this study. The objective of this study is to develop and optimize the features and content of the app with and for our target population. This methodology also has several strengths. In contrast to many other behavior change apps [65], we followed a user-centered design approach to optimize the app for our intended audience, which may have different content wishes than other populations [57,66,67]. We also followed app design best practices by conducting iterative rounds of testing, allowing ample opportunities for usability feedback from potential end users [40,42]. Finally, 2 team members experienced in qualitative research analyzed all participant feedback using more traditional qualitative content analysis techniques to search for generalizable insights that may inform future health intervention research beyond the development of this app.

Conclusions

Cancer prevention in the modern era must include options that are accessible to all, but this does not mean that all options must be universal. A mobile app—or any intervention, importantly—that promotes healthy, cancer-preventing behaviors in population A is not guaranteed to work as well with population B. Accordingly, given the disproportionate burden of cancer and cancer-related mortality among non-Hispanic Black populations in the United States [6-9], our iterative development approach for a cancer prevention mobile app focused uniquely and specifically on goal setting among non-Hispanic Black primary care patients. This iterative process led to the development of a cancer prevention mobile app that potential end users deemed usable and acceptable and yielded noteworthy insights about what intended end users value in health goals and how they may work on these goals with others.

Conflicts of Interest

LJ was employed by the app developer Transmogrify. Transmogrify had no input in the study design or evaluation.

Multimedia Appendix 1

Interview scripts and early prototype examples.

[DOCX File , 18180 KB - formative_v6i3e28157_app1.docx]

Multimedia Appendix 2

Interactive prototype and corresponding interview guide.

[DOCX File , 17 KB - formative_v6i3e28157_app2.docx]

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA A Cancer J Clin 2019;69(1):7-34. [doi: 10.3322/caac.21551]
2. Katzke VA, Kaaks R, Kühn T. Lifestyle and cancer risk. Cancer J 2015;21(2):104-110. [doi: 10.1097/PPO.0000000000000101] [Medline: 25815850]
3. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K, International Agency for Research on Cancer Handbook Working Group. Body fatness and cancer--viewpoint of the IARC working group. N Engl J Med 2016;375(8):794-798 [FREE Full text] [doi: 10.1056/NEJMsr1606602] [Medline: 27557308]
4. Anand P, Kunnumakkara AB, Sundaram C, Harikumar KB, Tharakan ST, Lai OS, et al. Cancer is a preventable disease that requires major lifestyle changes. Pharm Res 2008;25(9):2097-2116 [FREE Full text] [doi: 10.1007/s11095-008-9661-9] [Medline: 18626751]
5. Lortet-Tieulent J, Goding Sauer A, Siegel RL, Miller KD, Islami F, Fedewa SA, et al. State-level cancer mortality attributable to cigarette smoking in the United States. JAMA Intern Med 2016;176(12):1792-1798. [doi: 10.1001/jamainternmed.2016.6530] [Medline: 27775761]
6. Barbeau EM, Krieger N, Soobader MJ. Working class matters: socioeconomic disadvantage, race/ethnicity, gender, and smoking in NHIS 2000. Am J Public Health 2004;94(2):269-278. [doi: 10.2105/ajph.94.2.269] [Medline: 14759942]
7. Krieger N. Defining and investigating social disparities in cancer: critical issues. Cancer Causes Control 2005;16(1):5-14. [doi: 10.1007/s10552-004-1251-5] [Medline: 15750853]
8. Kawachi I, Berkman LF. Social capital, social cohesion, and health Internet. In: Berkman LF, Kawachi I, Glymour MM, editors. Social epidemiology. 2nd edition. Oxford, UK: Oxford University Press; 2014:290-319.
9. Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. Am J Epidemiol 1979;109(2):186-204. [doi: 10.1093/oxfordjournals.aje.a112674] [Medline: 425958]
10. Aysola J, Bitton A, Zaslavsky AM, Ayanian JZ. Quality and equity of primary care with patient-centered medical homes: results from a national survey. Med Care 2013;51(1):68-77 [FREE Full text] [doi: 10.1097/MLR.0b013e318270bb0d] [Medline: 23047125]
11. Bach PB, Pham HH, Schrag D, Tate RC, Hargraves JL. Primary care physicians who treat blacks and whites. N Engl J Med 2004;351(6):575-584. [doi: 10.1056/NEJMsa040609] [Medline: 15295050]
12. Cornwell B, Schumm LP, Laumann EO, Graber J. Social networks in the NSHAP study: rationale, measurement, and preliminary findings. J Gerontol B Psychol Sci Soc Sci 2009;64(Suppl 1):i47-i55 [FREE Full text] [doi: 10.1093/geronb/gbp042] [Medline: 19502574]
13. Freeman HP. Poverty, culture, and social injustice: determinants of cancer disparities. CA Cancer J Clin 2004;54(2):72-77 [FREE Full text] [doi: 10.3322/canjclin.54.2.72] [Medline: 15061597]
14. Haines VA, Hurlbert JS, Beggs JJ. Exploring the determinants of support provision: provider characteristics, personal networks, community contexts, and support following life events. J Health Soc Behav 1996;37(3):252-264. [Medline: 8898496]
15. Katz SJ, Hofer TP. Socioeconomic disparities in preventive care persist despite universal coverage. Breast and cervical cancer screening in Ontario and the United States. JAMA 1994;272(7):530-534. [Medline: 8046807]
16. Institute of Medicine (US) Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care. In: Smedley BD, Stith AY, Nelson AR, editors. Unequal treatment: confronting racial and ethnic disparities in health care. Washington, D.C: National Academies Press (US); 2003.
17. Gotay CC. Behavior and cancer prevention. J Clin Oncol 2005;23(2):301-310. [doi: 10.1200/JCO.2005.06.069] [Medline: 15637393]
18. Strecher VJ, Seijts GH, Kok GJ, Latham GP, Glasgow R, DeVellis B, et al. Goal setting as a strategy for health behavior change. Health Educ Q 1995;22(2):190-200. [doi: 10.1177/109019819502200207] [Medline: 7622387]
19. Bovend'Eerdt TJ, Botell RE, Wade DT. Writing SMART rehabilitation goals and achieving goal attainment scaling: a practical guide. Clin Rehabil 2009;23(4):352-361. [doi: 10.1177/0269215508101741] [Medline: 19237435]
20. Schunk DH. Goal setting and self-efficacy during self-regulated learning. Educ Psychol 1990;25(1):71-86. [doi: 10.1207/s15326985ep2501_6]
21. Heath C, Larrick RP, Wu G. Goals as reference points. Cogn Psychol 1999;38(1):79-109. [doi: 10.1006/cogp.1998.0708] [Medline: 10090799]

22. Bandura A. Health promotion from the perspective of social cognitive theory. Psychol Health 1998;13(4):623-649. [doi: 10.1080/08870449808407422]

23. O'Malley AJ, Christakis NA. Longitudinal analysis of large social networks: estimating the effect of health traits on changes in friendship ties. Stat Med 2011;30(9):950-964 [FREE Full text] [doi: 10.1002/sim.4190] [Medline: 21287589]

24. Centola D. An experimental study of homophily in the adoption of health behavior. Science 2011;334(6060):1269-1272. [doi: 10.1126/science.1207055] [Medline: 22144624]

25. Perkins JM, Subramanian SV, Christakis NA. Social networks and health: a systematic review of sociocentric network studies in low- and middle-income countries. Soc Sci Med 2015;125:60-78 [FREE Full text] [doi: 10.1016/j.socscimed.2014.08.019] [Medline: 25442969]

26. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med 2008;358(21):2249-2258 [FREE Full text] [doi: 10.1056/NEJMsa0706154] [Medline: 18499567]

27. Centola D. The spread of behavior in an online social network experiment. Science 2010;329(5996):1194-1197. [doi: 10.1126/science.1185231] [Medline: 20813952]

28. Wolff M, Bates T, Beck B, Young S, Ahmed SM, Maurana C. Cancer prevention in underserved African American communities: barriers and effective strategies--a review of the literature. WMJ 2003;102(5):36-40. [Medline: 14621929]

29. Thomson CA, McCullough ML, Wertheim BC, Chlebowski RT, Martinez ME, Stefanick ML, et al. Nutrition and physical activity cancer prevention guidelines, cancer risk, and mortality in the women's health initiative. Cancer Prev Res (Phila) 2014;7(1):42-53 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-13-0258] [Medline: 24403289]

30. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]

31. Chang E, Blondon K, Lyles CR, Jordan L, Ralston JD. Racial/ethnic variation in devices used to access patient portals. Am J Manag Care 2018;24(1):e1-e8 [FREE Full text] [Medline: 29350513]

32. Ray R, Sewell AA, Gilbert KL, Roberts JD. Missed opportunity? Leveraging mobile technology to reduce racial health disparities. J Health Polit Policy Law 2017;42(5):901-924. [doi: 10.1215/03616878-3940477] [Medline: 28663182]

33. Bennett GG, Steinberg DM, Stoute C, Lanpher M, Lane I, Askew S, et al. Electronic health (eHealth) interventions for weight management among racial/ethnic minority adults: a systematic review. Obes Rev 2014;15(Suppl 4):146-158. [doi: 10.1111/obr.12218] [Medline: 25196411]

34. James DC, Harville 2nd C, Sears C, Efunbumi O, Bondoc I. Participation of African Americans in e-Health and m-Health studies: a systematic review. Telemed J E Health 2017;23(5):351-364. [doi: 10.1089/tmj.2016.0067] [Medline: 27792475]

35. AlJaberi H. Developing culturally sensitive mHealth apps for Caribbean immigrant women to use during pregnancy: focus group study. JMIR Hum Factors 2018;5(4):e29 [FREE Full text] [doi: 10.2196/humanfactors.9787] [Medline: 30305256]

36. Alsswey AH, Al-Samarraie H, El-Qirem F, Alzahrani AI, Alfarraj O. Culture in the design of mHealth UI: an effort to increase acceptance among culturally specific groups. Electron Libr 2020;38(2):257-272. [doi: 10.1108/el-04-2019-0097]

37. Joseph RP, Keller C, Vega-López S, Adams MA, English R, Hollingshead K, et al. A culturally relevant smartphone-delivered physical activity intervention for African American women: development and initial usability tests of smart walk. JMIR Mhealth Uhealth 2020;8(3):e15346 [FREE Full text] [doi: 10.2196/15346] [Medline: 32130198]

38. Vardeman-Winter J. The framing of women and health disparities: a critical look at race, gender, and class from the perspectives of grassroots health communicators. Health Commun 2017;32(5):629-638. [doi: 10.1080/10410236.2016.1160318] [Medline: 27367141]

39. Vardeman-Winter J. Confronting whiteness in public relations campaigns and research with women. J Public Relat Res 2011;23(4):412-441. [doi: 10.1080/1062726x.2011.605973]

40. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]

41. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. In: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. 1993 Presented at: INTERCHI '93; April 24-29, 1993; Amsterdam, The Netherlands p. 206-213 URL: https://dl.acm.org/doi/10.1145/169059.169166 [doi: 10.1145/169059.169166]

42. Nielsen J. Estimating the number of subjects needed for a thinking aloud test. Int J Hum Comput Stud 1994;41(3):385-397. [doi: 10.1006/ijhc.1994.1065]

43. Virzi RA. Refining the test phase of usability evaluation: how many subjects is enough? Hum Factors 1992;34(4):457-468. [doi: 10.1177/001872089203400407]

44. Mason M. Sample size and saturation in PhD studies using qualitative interviews. Forum Qual 2010;11(3). [doi: 10.17169/fqs-11.3.1428]

45. O'Neill J, Conzemius AE. The power of SMART goals: using goals to improve student learning. Bloomington, IN: Solution Tree; 2005.

46. American cancer society guidelines for nutrition and physical activity. American Cancer Society. 2020. URL: https://www.cancer.org/healthy/eat-healthy-get-active/acs-guidelines-nutrition-physical-activity-cancer-prevention/guidelines.html [accessed 2021-02-11]

47. Si L, Callan J. A statistical model for scientific readability. In: Proceedings of the 10th International Conference on Information and knowledge management. New York: Association for Computing Machinery; 2001 Presented at: CIKM '01; October 5-10, 2001; Atlanta, GA p. 574-576. [doi: 10.1145/502585.502695]

48. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)– a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]

49. InVision. URL: https://www.invisionapp.com [accessed 2021-09-28]

50. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum-Comput Interact 2008;24(6):574-594. [doi: 10.1080/10447310802205776]

51. White MD, Marsh EE. Content analysis: a flexible methodology. Lib Trends 2006;55(1):22-45. [doi: 10.1353/lib.2006.0053]

52. Stemler SE. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. Pract Assess Res Eval 2019;9(1):4 [FREE Full text] [doi: 10.7275/96jp-xz07]

53. Resnick D, Schapira MM, Smith JM, Bautista A, Xu C, Jones L, et al. Promoting collaborative goal setting for cancer prevention among primary care patients through mHealth: mixed methods evaluation of a new app. JMIR Form Res 2021;5(7):e22510 [FREE Full text] [doi: 10.2196/22510] [Medline: 34259162]

54. Satia JA, Barlow J, Armstrong-Brown J, Watters JL. Qualitative study to explore prospect theory and message framing and diet and cancer prevention-related issues among African American adolescents. Cancer Nurs 2010;33(2):102-109 [FREE Full text] [doi: 10.1097/NCC.0b013e3181be5e8a] [Medline: 20142738]

55. Webb TL, Sheeran P. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychol Bull 2006;132(2):249-268. [doi: 10.1037/0033-2909.132.2.249] [Medline: 16536643]

56. Gallagher KM, Updegraff JA. Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. Ann Behav Med 2012;43(1):101-116. [doi: 10.1007/s12160-011-9308-7] [Medline: 21993844]

57. Tarver WL, Haggstrom DA. The use of cancer-specific patient-centered technologies among underserved populations in the United States: systematic review. J Med Internet Res 2019;21(4):e10256 [FREE Full text] [doi: 10.2196/10256] [Medline: 31012855]

58. Patel MS, Small DS, Harrison JD, Fortunato MP, Oon AL, Rareshide CA, et al. Effectiveness of behaviorally designed gamification interventions with social incentives for increasing physical activity among overweight and obese adults across the United States: the STEP UP randomized clinical trial. JAMA Intern Med 2019;179(12):1624-1632 [FREE Full text] [doi: 10.1001/jamainternmed.2019.3505] [Medline: 31498375]

59. Brewer LC, Kumbamu A, Smith C, Jenkins S, Jones C, Hayes SN, et al. A cardiovascular health and wellness mobile health intervention among church-going African Americans: formative evaluation of the FAITH! app. JMIR Form Res 2020;4(11):e21450 [FREE Full text] [doi: 10.2196/21450] [Medline: 33200999]

60. Coorey GM, Neubeck L, Mulley J, Redfern J. Effectiveness, acceptability and usefulness of mobile applications for cardiovascular disease self-management: systematic review with meta-synthesis of quantitative and qualitative data. Eur J Prev Cardiol 2018;25(5):505-521. [doi: 10.1177/2047487317750913] [Medline: 29313363]

61. Dugas M, Gao GG, Agarwal R. Unpacking mHealth interventions: a systematic review of behavior change techniques used in randomized controlled trials assessing mHealth effectiveness. Digit Health 2020;6:2055207620905411 [FREE Full text] [doi: 10.1177/2055207620905411] [Medline: 32128233]

62. Peek ME, Sayad JV, Markwardt R. Fear, fatalism and breast cancer screening in low-income African-American women: the role of clinicians and the health care system. J Gen Intern Med 2008;23(11):1847-1853 [FREE Full text] [doi: 10.1007/s11606-008-0756-0] [Medline: 18751758]

63. Flórez KR, Aguirre AN, Viladrich A, Céspedes A, De La Cruz AA, Abraído-Lanza AF. Fatalism or destiny? A qualitative study and interpretative framework on Dominican women's breast cancer beliefs. J Immigr Minor Health 2009;11(4):291-301 [FREE Full text] [doi: 10.1007/s10903-008-9118-6] [Medline: 18253833]

64. Spurlock WR, Cullins LS. Cancer fatalism and breast cancer screening in African American women. ABNF J 2006;17(1):38-43. [Medline: 16596899]

65. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of publicly available physical activity apps: review and content analysis. JMIR Mhealth Uhealth 2018;6(3):e53 [FREE Full text] [doi: 10.2196/mhealth.9069] [Medline: 29563080]

66. Brewer LC, Fortuna KL, Jones C, Walker R, Hayes SN, Patten CA, et al. Back to the future: achieving health equity through health informatics and digital health. JMIR Mhealth Uhealth 2020;8(1):e14512 [FREE Full text] [doi: 10.2196/14512] [Medline: 31934874]

67. Armaou M, Araviaki E, Musikanski L. eHealth and mHealth interventions for ethnic minority and historically underserved populations in developed countries: an umbrella review. Int J Commun Wellbeing 2019;3(2):193-221. [doi: 10.1007/s42413-019-00055-5]

Abbreviations

ACC: Abramson Cancer Center

ACS: American Cancer Society
BIPOC: Black, Indigenous, and People of Color
PCP: primary care provider
REDCap: Research Electronic Data Capture
SMART: specific, measurable, achievable, realistic, and time-bound

A Novel Diagnostic Decision Support System for Medical Professionals: Prospective Feasibility Study

Joanna Timiliotis[1*], MSc; Bibiana Blümke[1,2*], Dr Med; Peter Daniel Serfözö[1], Dr Med; Stephen Gilbert[2,3], PhD; Marta Ondrésik[2], PhD; Ewelina Türk[2], MD; Martin Christian Hirsch[2,4], Prof Dr; Jens Eckstein[1,5], Prof Dr, PhD

[1]CMIO Research Group, Digitalization & ICT Department, University Hospital Basel, Basel, Switzerland

[2]Ada Health GmbH, Berlin, Germany

[3]Else Kröner Fresenius Center for Digital Health, University Hospital Carl Gustav Carus Dresden, Technische Universität Dresden, Dresden, Germany

[4]Institute for Artificial Intelligence in Medicine, Philipps University of Marburg, Marburg, Germany

[5]Department of Internal Medicine, University Hospital Basel, Basel, Switzerland

*these authors contributed equally

Corresponding Author:
Joanna Timiliotis, MSc
CMIO Research Group
Digitalization & ICT Department
University Hospital Basel
Hebelstrasse, 10
Basel, 4031
Switzerland
Phone: 41 0613285489
Email: joanna.timiliotis@usb.ch

Abstract

Background: Continuously growing medical knowledge and the increasing amount of data make it difficult for medical professionals to keep track of all new information and to place it in the context of existing information. A variety of digital technologies and artificial intelligence–based methods are currently available as persuasive tools to empower physicians in clinical decision-making and improve health care quality. A novel diagnostic decision support system (DDSS) prototype developed by Ada Health GmbH with a focus on traceability, transparency, and usability will be examined more closely in this study.

Objective: The aim of this study is to test the feasibility and functionality of a novel DDSS prototype, exploring its potential and performance in identifying the underlying cause of acute dyspnea in patients at the University Hospital Basel.

Methods: A prospective, observational feasibility study was conducted at the emergency department (ED) and internal medicine ward of the University Hospital Basel, Switzerland. A convenience sample of 20 adult patients admitted to the ED with dyspnea as the chief complaint and a high probability of inpatient admission was selected. A study physician followed the patients admitted to the ED throughout the hospitalization without interfering with the routine clinical work. Routinely collected health-related personal data from these patients were entered into the DDSS prototype. The DDSS prototype's resulting disease probability list was compared with the gold-standard main diagnosis provided by the treating physician.

Results: The DDSS presented information with high clarity and had a user-friendly, novel, and transparent interface. The DDSS prototype was not perfectly suited for the ED as case entry was time-consuming (1.5-2 hours per case). It provided accurate decision support in the clinical inpatient setting (average of cases in which the correct diagnosis was the first diagnosis listed: 6/20, 30%, SD 2.10%; average of cases in which the correct diagnosis was listed as one of the top 3: 11/20, 55%, SD 2.39%; average of cases in which the correct diagnosis was listed as one of the top 5: 14/20, 70%, SD 2.26%) in patients with dyspnea as the main presenting complaint.

Conclusions: The study of the feasibility and functionality of the tool was successful, with some limitations. Used in the right place, the DDSS has the potential to support physicians in their decision-making process by showing new pathways and unintentionally ignored diagnoses. The DDSS prototype had some limitations regarding the process of data input, diagnostic accuracy, and completeness of the integrated medical knowledge. The results of this study provide a basis for the tool's further development. In addition, future studies should be conducted with the aim to overcome the current limitations of the tool and study design.

XSL·FO
RenderX

KEYWORDS

diagnostic decision support system; DDSS; probabilistic reasoning; artificial intelligence; dyspnea; emergency department; internal medicine; symptom checker

Introduction

Background

Digital tools play an increasingly relevant role in the health sector. Most patients search the internet to complement their knowledge of health care topics [1]. In addition, patients increasingly use *symptom checkers* instead of standard search engines for symptom analysis [2]. In contrast to the fast uptake in the consumer sector, professional tools similar to symptom checkers designed to support physician decision-making have not found widespread adoption in the clinical and outpatient environment [3] even though this concept is not new. A variety of digital technologies and artificial intelligence–based methods are currently available and have recently emerged as impressively persuasive tools to empower physicians in clinical decision-making and improve health care quality [4]. Diagnostic decision support systems (DDSSs) have been demonstrated to facilitate the assessment of clinical data input by using an extensive medical knowledge base [5,6].

Continuously growing medical knowledge and the increasing amount of data make it difficult for medical professionals to keep track of all new information and to place it in the context of existing information [7]. DDSSs have been suggested as a solution to this problem [8]. An expert system can help by expanding the clinician's differential diagnosis list and suggesting other avenues of investigation [9].

Diagnostic errors, consisting of inaccurate, delayed, or missed diagnoses, remain major challenges in public health care [10] that need to be addressed.

The overall purpose is to invite physicians to rethink and re-examine their steps and possible alternatives in light of the presented diagnostic information [11]. DDSSs are not intended to replace physicians but rather to augment and optimize the diagnostic decision-making process. If they are to be adopted, it is important that they provide accurate information and are trusted by clinicians. The diagnostic decision-making process must be as transparent and comprehensible as possible. The trustworthiness of the data handling and the medical quality of the knowledge base and algorithms are essential to this. Poor usability is another important barrier that could limit adoption and be a deterrent to the routine use of new technology.

In this study, we pilot-tested whether the use of a DDSS prototype from Ada Health GmbH is feasible in an emergency department (ED) setting.

The Diagnostic Decision Support Tool

The DDSS is a web-based diagnostic decision support system for medical professionals developed as a research prototype by Ada Health GmbH that can be accessed by laptop or tablet.

In the DDSS, the physician can input a patient case over time with several visits (if relevant), and the system updates the provided decision support dynamically. The user interface consists of pages representing the steps during an individual patient visit and a case overview page. The design of this prototype provides full transparency over the artificial intelligence–based medical reasoning. The user interface allows for a continuous and transparent exchange between the machine and human.

The case starts with the input of the epidemiological data followed by a consultation page where one or several findings are entered. On the case analysis dashboard, symptoms, findings, and their related attributes can be added as *present* or *absent* for the case. The search allows the user to enter synonyms or related terms to find a specific symptom. In addition, the tool suggests a ranked list that changes in real time of symptoms and findings that have the highest potential for information gain for the current case. The DDSS supports the collection of both patient-reported complaints and findings gathered via medical examination or testing. Lifestyle or risk factors that may affect the patient's condition can also be recorded, and it is taken into consideration via the reasoning engine. The system does not use a predefined standard ontology or taxonomy to enter symptoms.

The patient information, symptoms, and findings, as well as a list of differential diagnoses ranked by *probability* and *fit*, are represented on the main page of the tool (Figure 1).

Figure 1. Screenshot of the case analysis dashboard of the diagnostic decision support system prototype.
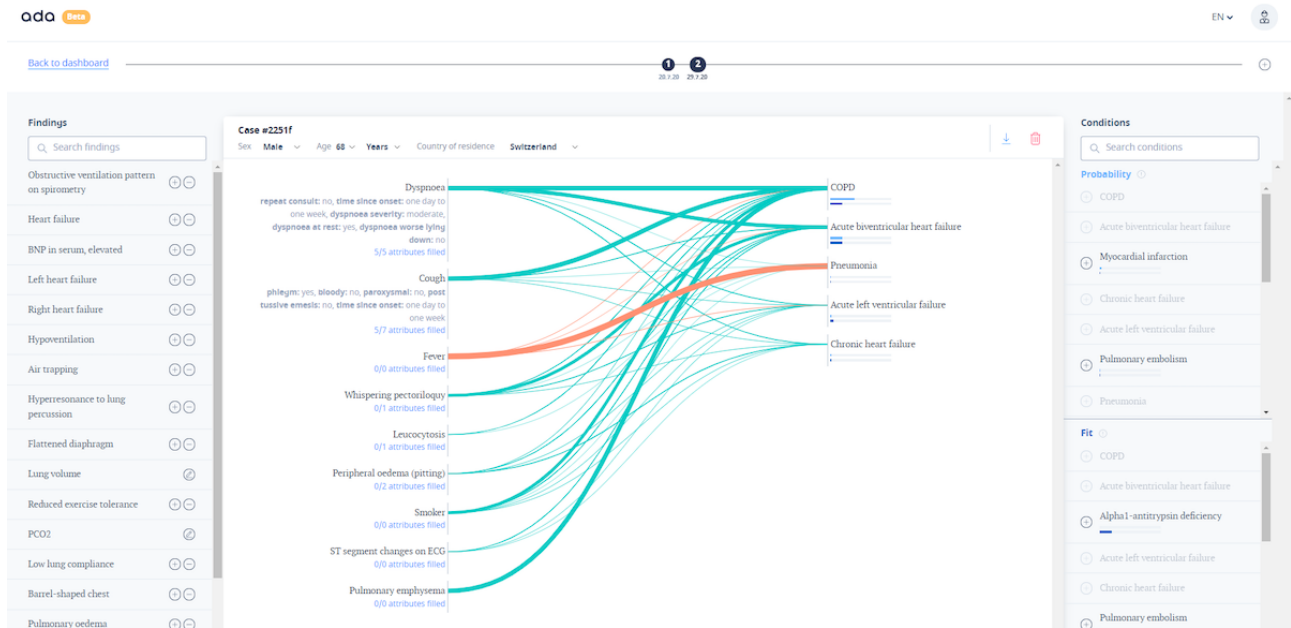
The *probability* list is ranked by the estimated probability of a disease. It is based on the representation of medical knowledge using a probabilistic reasoning engine considering existing epidemiological data such as age, sex, or geographical location. This mirrors the approach that a health care professional takes during clinical routine. The *fit* list is ranked by the most likely conditions that could explain the finding constellation without knowledge of the probability of the conditions occurring in the general population. The reasoning engine infers disease probability estimations based on a representation of medical knowledge. The medical knowledge base is used to define a Bayesian network in which approximate inference is carried out.

Contribution lines visualize the correspondence between a symptom and a disease. The relative weighting of the symptoms to the diseases is indicated by the thickness of the lines. The color of the lines indicates the presence or absence of the finding in the constellation. This user interface was designed to ensure the transparency of the underlying reasoning engine inferences to the physician in real time. The medical knowledge base of the prototype DDSS was not based on a pre-existing database or medical knowledge ontology. Instead, it was generated and reviewed by in-house medical professionals using a process of curated integration of peer-reviewed medical literature. The medical knowledge and reasoning of the tool were designed with the primary goal of achieving high condition suggestion accuracy. More than 1300 conditions and 11,000 findings and symptoms are available in the medical knowledge base.

The DDSS prototype has been examined in a retrospective study with a focus on rare diseases, demonstrating that Ada suggested accurate diagnoses earlier than clinical diagnoses in more than half of all cases [12]. However, this tool has not been investigated prospectively in a real-life setting. The DDSS is a prototype in development, has not yet been optimized for everyday use, and is not publicly available. Nevertheless, the user interface is novel and unique in its presentation and

transparency. In this regard, we aimed to conduct feasibility testing with a focus on a very common symptom; namely, *dyspnea*. Patients presenting to the ED with dyspnea were chosen as the focus area for testing as dyspnea has a wide range of possible etiologies, including cardiac, pulmonary, and infectious diseases [13]. This approach ensured a broad range of possible outcomes to comprehensively test the system's novel user interface while being appropriate to the stage of development of the prototype.

Aim of the Study

This is the first prospective study evaluating a DDSS prototype from Ada Health GmbH, which uses a novel approach for dynamically interacting with the physician in a real-life clinical setting by entering routinely collected health-related personal data. Our primary goal is to investigate the potential of this concept. Secondary outcomes are the identification of any key reasons for inaccuracy, current technical limitations, and the potential for further development and adaptation of the DDSS prototype based on the findings and needs identified with regard to the usability of the tool.

Methods

Study Design and Case Selection

We conducted this prospective feasibility study (ClinicalTrials.gov: NCT0482734) at the ED and internal medicine ward of the University Hospital Basel, Switzerland. A convenience sample of 20 adult patients admitted to the ED with a chief complaint of dyspnea and a high likelihood of inpatient admission was selected. The participants had to be able to understand, speak, and read in German. The exclusion criteria were refusal of consent and discharge from the ED without inpatient admission. The study period was from May 2020 to August 2020. The study participation of each patient lasted as long as the patient stayed in the hospital.

This study design was observational—patients with dyspnea admitted to the ED of the University Hospital Basel were monitored, diagnosed, and treated according to the usual clinical routine. The study physician (PDS) shadowed the treating physician and the patient throughout the entire hospitalization without any interference with the routine clinical work (Figure 2).

Figure 2. Study process. CIS: clinical information system; DDSS: diagnostic decision support system; ED: emergency department.
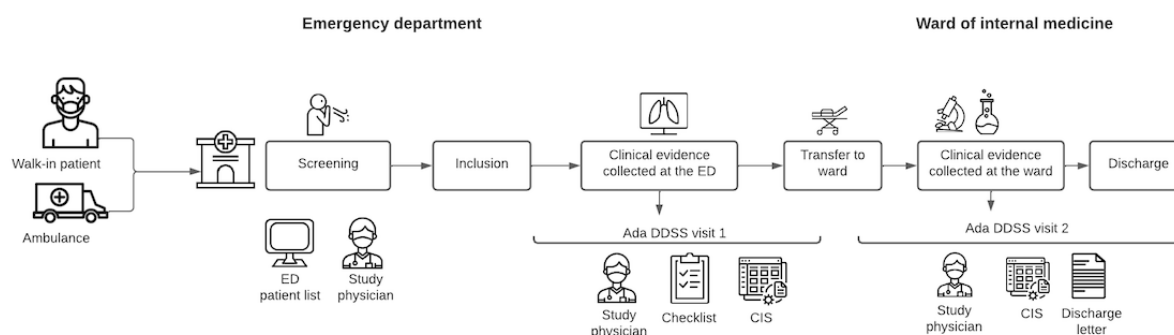
A first evaluation of potential patients for recruitment against the inclusion and exclusion criteria was based on a patient's medical file using the triage findings from the ED. The decision whether the patient could be included was made after the first patient contact (ie, within hours of the patient's arrival at the ED). The investigator explained the objective of the study and its observational nature to the patient. For ethical and organizational reasons, consent from the patient was obtained post hoc once the patient was hospitalized (following Human Research Act Article 31 [14] and Clinical Trials Ordinance Article 15 [15]). Data from patients who refused to provide post hoc consent were no longer used for the research project.

Data Acquisition

Data Collection

All patients underwent an initial clinical assessment at the ED in which the study physician used a checklist to document symptoms, medical history, vital signs, and physical examination in a structured manner. Complementary information documented by the treating care team during hospitalization as well as all other investigation findings were extracted from the medical record.

The treating physician at no point had access to insights into the case from the DDSS prototype. The patients received usual care from their examining and treating medical staff. The study investigator (PDS) was not involved in patient care at any point.

Prototype DDSS Input

Once a patient was admitted, a new case was created in the DDSS with the patient's sex, age, and geographical location as the first information. As we focused on patients with dyspnea in this study, a new patient case was started by entering the finding *Dyspnea* and selecting the corresponding attributes and specifications. All clinical evidence collected from the patient was entered as DDSS input data to build the case (Figure 1). Findings that would have been marked as *absent* (eg, no fever) were only added if relevant to the list of diagnoses or if explicitly mentioned in the medical record.

Information from the medical record was assigned to the time of the visit in the DDSS prototype. The idea was to mirror the patient's journey in the hospital and provide the system with the same amount of information the treating physician had at a certain point in time. The first visit (visit 1) in the DDSS was created at the end of the ED stay. All evidence prospectively collected until this moment was entered into the DDSS. The second visit (visit 2) in the DDSS corresponded to when the patient was discharged from the hospital. All information from the first visit was transferred to the second visit, modified if necessary, and complemented with additional information from the patient file gained during the hospital stay. Any information of potential relevance to diagnosis that could not be entered into the DDSS as it was not found in the tool was recorded in a separate document. Missing diagnoses were also noted.

An additional visit (visit 2.1) in the Ada DDSS was performed retrospectively by a physician and former Ada employee with expert knowledge of the medical content and technical aspects of the Ada DDSS. This person screened the clinical cases and lists of missing information in the tool generated by the study research team. The goal was to show the user dependency of the DDSS and the influence of this on the accuracy of the DDSS suggestions.

All inputs were performed in German as all clinical evidence was gathered in German.

Feasibility and Usability of the DDSS

The time of data entry, search functions and functionalities, availability of findings and diagnoses, and applicability of the tool in an acute ED setting were recorded to assess the feasibility of the DDSS prototype. We also evaluated the workflow and whether the navigation, data entry, and retrieval would impede clinical task completion. Furthermore, the input procedure with the tool's robustness to irrelevant variations in input data as well as the technical aspects and potential restrictions were analyzed.

The usability was assessed by considering the structure and composition of the DDSS interface and whether it was satisfactory.

As the novelty of the tool is mainly reflected in the design of the interface, this was a key object of investigation. Therefore,

the clarity of the visual representation of clinical data and the ease of acquiring information at a glance were examined in detail.

The guidance through different levels of the tool (onboarding, consultation page, and case analysis) was another point of interest. We tested whether the logic and availability of the desired options were consistent and rigorous and if the tool provided an effective layout. A risk evaluation of the misinterpretation of information was conducted.

Metrics

Different metrics for the assessment of the accuracy of the DDSS suggestions are listed and defined in Textbox 1.

Textbox 1. Metrics for the assessment of the accuracy of the diagnostic decision support system suggestions [5].

- Correct or accurate diagnosis retrieval: proportion of cases in which the correct diagnosis was the first diagnosis listed (M1), listed as one of the top 3 (M3), or listed as one of the top 5 (M5)

- Diagnosis in knowledge base: proportion of the diagnoses that were included in the knowledge base of the tool

---

The accuracy of the tool's diagnostic suggestions was evaluated at each of the different time points of the hospital stay. The list of the 5 most probable conditions provided by the tool was recorded for visit 1, visit 2, and visit 2.1. For the same time points, a maximum of 5 diagnoses were provided in the medical record. If <5 diagnoses were provided in the hospital, the total number of condition suggestions for the case from the DDSS was reduced accordingly. The first listed diagnosis in the ED and the final discharge diagnosis from the hospital were defined as the gold-standard diagnosis for visits 1, 2, and 2.1. The top 1 diagnosis, the top 3 diagnoses, and the top 5 diagnoses provided by the DDSS disease *probability* list were compared with this gold-standard main diagnosis for the 3 different visits. Furthermore, the proportion of diagnoses included in the knowledge base of the tool was assessed. In addition, the missing potentially important information of the findings and symptoms for each case was analyzed and categorized.

Matching of Diagnoses

The first 5 diagnoses from the DDSS and the first 5 diagnoses from the medical record for each visit were shown to 3 different physicians separately and independently following the completed data collection. They decided whether the diagnosis from the DDSS matched the diagnosis from the treating physician at the different time stamps. This process was necessary as the naming and, therefore, the interpretation of the matching of the diagnoses were not standardized. The 3 physicians did not see the entire case, only the diagnosis lists. They had different levels of clinical expertise and knowledge of the DDSS and the patient case. None of the physicians were involved in anamnesis, clinical examination, or treatment of the patient. The three physicians comprised (1) the study physician, who was involved in the data collection and entry into the DDSS prototype and was therefore familiar with the patient case and who saw the patient in the ED to obtain informed consent and evaluate the appropriateness of the patient for study inclusion; (2) a second independent physician who was an experienced senior physician and fellow of internal medicine and cardiology; and (3) a third physician with several years of work experience in the Medical Knowledge Team of Ada Health GmbH in Berlin and detailed knowledge of the available medical content and the reasoning engine of the DDSS and who also saw the DDSS case in detail to analyze potential user dependencies.

Precise matching criteria were not specified; instead, the physicians were directed to use their experience to decide on the appropriateness of the DDSS condition suggestions. The reasons for inaccurate suggestions of diagnoses were analyzed and categorized. Several causes per case could be assigned, and the related proportions were calculated.

Statistical Analysis

The top-1, top-3, and top-5 performance of the Ada DDSS prototype condition suggestion for each of the visits, with comparison against the ratings by the 3 physicians, were compared using descriptive statistics and tests appropriate for categorical data. Chi-square tests were used to test whether the proportion of correct answers was drawn from the same distribution, with the application of this test across all visits, once for each of the metrics for comparison (top 1, top 3, and top 5 matching condition suggestions) for each of the 3 physicians' ratings, to be followed in case of a significant difference by post hoc 2-sided pairwise Fisher exact tests [16]. *P* values were corrected for multiple comparisons using the Benjamini–Hochberg procedure [17], guided by the interpretation of Armstrong [18], and considered significant if <.05.

Data Processing and Ethical Approval

The conducted study complied with the ethical principles of the World Medical Association Declaration of Helsinki [19]. Ethical approval was obtained from the Ethics Commission Nord-West-Schweiz (reference 2020-00095, date of approval January 24, 2020). All data were stored and transferred in a pseudonymized form. Data processing and transfer were performed in accordance with national and local guidelines. An order data processing agreement was made between the University Hospital Basel and Ada Health GmbH.

Results

Patient Characteristics

A total of 33 patients with dyspnea were considered for inclusion, of which 61% (20/33) cases were included and 33% (11/33) were excluded because of direct discharge from the ED without referral to the ward. The refusal rate was low (2/33, 6%). The resulting study population consisted of 40% (8/20) women and 60% (12/20) men aged 54 to 93 years (mean 74 years, SD 10.44 years).

Feasibility Measures

To create a case in the DDSS prototype, it is required to enter basically all recorded patient data, which is time-consuming, especially in a setting as time-limited as the ED. The data entry took 1.5-2 hours per patient. The checklist used in the ED by the study physician resulted from a pilot round that preceded this study and was a key component of the initial feasibility findings. It was accepted at the outset of the study that the novel user interface of the DDSS prototype was not yet fully developed for use in parallel to every patient's examination or consultation. The checklist, observational recording, and non–real-time use of the system allowed for the identification of how such a prototype DDSS would need to be involved in capturing the high speed and complexity of clinical data delivery.

Usability

Overview

The Ada DDSS could be usable in the research setting; however, the research team considered that it required optimization before it could be adopted in everyday use in the ED.

Usability insights from our study were principally related to the DDSS *main page* (ie, the *Case analysis* page). This interface consists of three sections: findings and symptoms on the left, case dashboard in the middle, and diagnosis suggestions on the right side (Figure 1).

Usability of the Findings Section

The findings were easily located in the search function via several synonyms and terms by the study physician. However, the search engine contained some subcategories and finding synonyms that were sometimes misleading. For more specific findings such as *orthopnea*, the superordinate category *dyspnea* must first be selected and provided with corresponding attributes (in this case, *occurs while lying flat*). This led to a time-consuming search for the right designations by the study physician.

During case input in this study, the finding suggestions were rarely used as the ranking by probability often did not match the physician's natural clinical workflow.

Dashboard Usability

If a finding was added via search function or the list of relevant suggestions (left panel in Figure 1), it must be declared as *present* or *absent* before it was transferred to the case dashboard. Once added to the case, it was not immediately recognizable to the physician how the finding had been marked, which confused the study physician during his work. The color of the contribution line indicates the presence or absence of the finding, but the finding itself is not marked in either way.

The clarity of the presentation of the symptom constellation in relation to the diagnosis list by the contribution lines creates transparency for the user on how the reasoning engine is working. This is one of the main advantages of this DDSS prototype in comparison with others according to the study team.

Usability of the Diagnostic Suggestions

As soon as the first finding was entered, the 5 most probable suggested diagnoses were automatically transferred to the case dashboard when switching from the consultation page to the main page.

Once they were listed on the case dashboard, the diagnoses did not update themselves automatically when adding or deleting information. The *probability* and *fit* lists of potential diagnoses on the right side, in contrast, changed in real time, which was confusing for the study personnel.

Accuracy of Suggested Diagnoses

Overview

The results for the accuracy of the DDSS suggested diagnoses are shown in Table 1.

Table 1. Accuracy of suggested diagnoses compared with the gold-standard diagnosis (N=20).

	M1[a]			M2[b]			M3[c]		
	(%)	(n/N)	(SD)	(%)	(n/N)	(SD)	(%)	(n/N)	(SD)
Visit 1	18	3.67/20	0.58	42	8.33/20	1.53	57	11.33/20	0.58
Visit 2	35	7/20	1.00	62	12.33/20	1.53	75	15/20	1.00
Visit 2.1	40	8/20	1.00	62	12.33/20	1.53	80	16/20	1.00

[a]First diagnosis listed.

[b]One of the top 3 diagnoses listed.

[c]One of the top 5 diagnoses listed.

The table shows the average top-1, top-3, and top-5 accuracy of the DDSS's suggestions compared with the gold-standard diagnosis at the different visits, with assessment of the matching of the diagnosis suggestions by physicians with different levels of clinical and tool experience.

Different reasons for incorrect suggestion at the time of diagnosis could be identified and are listed in the following sections.

Multimorbidity or Multiple Confirmed Diagnoses or Symptom as Diagnosis

Inpatients in the department of internal medicine often have >1 diagnosed disease, either as known pre-existing diseases or as

unknown diseases diagnosed during the inpatient stay. The DDSS prototype seems to focus its reasoning on the evaluation of 1 main diagnosis and, thus, multimorbidity seems to be one of the biggest challenges in correcting condition suggestions. In a few cases, this was the main reason for an incorrect diagnostic suggestion in the DDSS.

In half of the cases (10/20, 50%), the treating physician did not provide a working or final diagnosis compliant with the International Classification of Diseases, 10th revision, as the first listed diagnosis but instead provided a list of several potential working diagnoses or a presenting complaint, which

made assessment of the accuracy of the DDSS suggestion impossible. This was especially true for visit 1, when the patient was transferred to the ward for further investigation.

Missing Entities in DDSS Knowledge Base

Another aspect that led to incorrect condition suggestions was the lack of relevant entities in the knowledge base of the tool, which limited its ability to suggest a diagnosis. In all these cases, there were one or more relevant diagnostic findings missing (Table 2). In addition, in 20% (4/20) of the cases, the final diagnosis did not exist in the DDSS knowledge base.

Table 2. Coverage of symptoms and findings in the medical knowledge base of the diagnostic decision support system (N=20).

Coverage category	Cases with missing relevant information, n (%)
Diagnostic findings (including laboratory, imaging, and histology)	20 (100)
Medical history (pre-existing condition, social, family, and medication)	20 (100)
History or examination findings	9 (45)
Attributes (investigation findings and factors)	6 (30)
Final diagnosis (first diagnosis)	4 (20)
Physiological finding≠negative pathological finding	4 (20)

User Input Dependencies

In a number of cases, the level of user experience with the tool was a decisive criterion for the subsequent accuracy of the diagnostic suggestions. In 10% (2/20) of the cases, it was essential for the physician inputting information into the DDSS to know the precise finding name in the DDSS to enable the system to provide accurate diagnosis suggestions.

Coverage of Diagnoses in the DDSS

For the analysis of the coverage of the clinical diagnoses in the knowledge base of the DDSS prototype (Table 3), a maximum of 5 confirmed diagnoses was considered. There was a total of 186 diagnoses for all cases. Each exact match was considered. In addition, each disease in the differential diagnosis list provided by the hospital was calculated as 0.5 if the diagnosis existed in the DDSS but not in the exact specification, grade, or localization described by the treating physician.

Table 3. Coverage of diagnoses in the knowledge base of the diagnostic decision support system.

Item	Visit 1	Visit 2
Sum of diagnoses	93	93
Sum of matches	56.5	65.5
Proportion (%; matches/diagnoses)	61	70

In all the cases, there were one or several entities (out of a large number of relevant entities for each case) that could not be found via the search function in the DDSS (Table 1). Those that could not be entered were mainly diagnostic findings, such as radiologic, laboratory, and histologic findings. In addition, the DDSS did not provide the possibility to report the medical history of pre-existing conditions, medications, or social and familial anamnesis.

Discussion

Principal Findings

Regarding the functionality and usability of the tool, it can be summarized that the dynamically interactive DDSS has high potential, with limitations. It showed convincing performance in its clarity of presentation (including transparency of the working of the medical engine) and provided a user-friendly interface. However, the tool as currently developed is not

perfectly suited to acute medical settings such as the ED as manual case entry is very time-consuming.

The findings on the DDSS disease suggestion accuracy indicate that it could provide accurate results in the clinical inpatient setting for the many patients who had dyspnea as the main presenting complaint. The symptom analysis algorithm of the DDSS seems to weigh the order of the symptoms present in a case, the likelihood of a finding for a diagnosis, and the epidemiology. Unlikely symptoms, absent common symptoms, and misleading findings as well as an atypical age of the patient for a disease or an uncommon primary anatomical site of involvement might lead to misdiagnosis in the system. These results should be interpreted with caution at this stage as the study setting was observational, and real-world interventional studies are suggested for confirmation.

Another finding of this study is that, although the medical professional knowledge base already covers many different

findings, it is nonetheless incomplete in some areas. Many findings from investigative procedures in the hospital are not yet provided by the tool and, in some cases, this decreased the accuracy of the suggested diagnoses.

The diagnosis suggestions also depend to a large degree on appropriate user input. The treating physician's medical knowledge and skills as well as the expertise of the study team with the use of the tool could potentially have influenced the outcome of this study. A higher experience in all of these fields might improve the accuracy outcome and should be investigated separately. It is acknowledged that the use of a DDSS of this type in a real-world setting requires training of personnel on the use of the system and how to obtain the best results from the tool.

The results from this study suggest that the Ada DDSS could have the potential to support the clinician in their daily work, but an enlargement of its professional medical knowledge base and a larger-scale evaluation study would be necessary beforehand.

Possible Improvements to the Ada DDSS

This feasibility study found some areas where the DDSS could be improved. The search and selection of symptoms and findings is one of the areas with the greatest potential for improvement. A structure that follows the logic of how physicians think (eg, a step-by-step selection starting with the examination or the investigation method, ending with the proofing pathological finding, and dividing the findings into categories) could improve the intuitive usability for physicians. The manner of displaying the highest–information-gain symptoms and finding suggestions on the case dashboard could thereby also be improved. In its current stage, this list shows a collection of unsorted and uncategorized symptoms and findings generated by the reasoning engine.

As it is up to the using physician to select the relevant diagnoses from the diagnosis lists and add them to the case dashboard, this should be made clear to the user, or the list of added diagnoses should automatically update itself. It would be helpful to signal to the user at a glance whether the finding was marked as absent either by using a different font color or by placing a cross in front of the finding. This would be a simple change with a large impact on usability.

The routine adoption of the tool in a highly dynamic setting such as the ED could only be achieved after a reduction in active effort to enter information. Automatic integration of basic patient and anamnestic information as well as further extraction of information from the electronic medical record could save a large amount of time for the treating physician and decrease potential bias because of user dependencies. The acceleration of rare disease diagnosis [12] and the higher accuracy in the inpatient setting also indicate that, in its current form, the DDSS is more suited to those disciplines. Used in the correct medical setting, it has the potential to support the physician in their decision-making process by showing new pathways of diagnostic reasoning and suggesting unintentionally ignored diagnoses. The pooling of the immense medical knowledge available has the potential to extend the medical disease spectrum of a physician in their routine work. If this functionality is extended through a wider professional medical knowledge base, it has the potential to assist in rational test choice and avoid important diagnostic investigations being overlooked.

Many patients in the internal medicine ward have >1 diagnosis. Multiple diagnoses at the time of discharge from the ED to the ward or from the ward to an outpatient setting led to a lower accuracy in the tool. This seems to be one of the biggest challenges and should be a focus area for the improvement of the DDSS.

Currently, medication and information about the therapy of a patient cannot be entered into the DDSS. This results in the underconsideration of possible therapeutic symptom improvements, therapy failure, or medicinal side effects in the probability estimation of the diagnosis suggestions.

Social or family history has a rudimentary representation in the DDSS and, consequently, follow-up or secondary diseases, exacerbation of an existing disease, social measures, or familial predisposition are underreflected.

An optimized and extended Ada DDSS based on the system evaluated here could save time and improve investigative and diagnostic efficiency and quality, thereby improving health economic outcomes [20,21]. These effects need to be assessed in future studies.

Strengths and Limitations

This study has several strengths and limitations. The study design, with a real-world setting, prospective data collection, and the shadowing of the treating physician by a study physician without any interference with usual care, as well as the measurement of accuracy analysis through the use of a panel of 3 physicians with different clinical backgrounds, was important for the study strengths.

The small number of cases is a limitation of the study, as is the focus on only 1 main presenting symptom for the selection of participants. Both factors offer potential for selection bias; however, they are appropriate for this stage of feasibility evaluation.

Our study showed that the DDSS condition suggestions are user-dependent—the level of knowledge, expertise, and familiarity had a large impact on accuracy. In some cases, an additional finding, which was difficult to find via the tool's search function because of specific wording, led to a completely altered differential diagnosis list and accuracy. As the case set was small and a large range of physician users was not explored, the range of user dependency of the DDSS was not precisely quantified. The focus of this study was to assess the feasibility and usability of the novel interface of the DDSS. It was not intended as a validation of the accuracy of its diagnosis suggestions. A future large-scale study with blinded, maybe automated data entry after consistent training of the clinical users of the system should be performed to evaluate and validate the accuracy of a ready-for-market DDSS.

XSL·FO

RenderX

Conclusions

This study provides insights into the applicability and performance of the DDSS prototype and the potential of the highly dynamic case input interface for medical professionals, especially in an inpatient setting. The clear and user-friendly presentation of a clinical patient case, with a transparent visual explanation of the algorithmic decisions, is the outstanding novelty of the tool used in this study.

At its current stage of development, the DDSS prototype has some limitations regarding the automation of data input, the accuracy of the diagnostic suggestions, and the completeness of the integrated medical knowledge. Data entry and analysis are still highly user-dependent; however, this could be minimized through training and experience.

The results of this study provide a basis for the further development of this and related tools. Further development of dynamic and highly transparent DDSS case interfaces is warranted and, once systems are optimized further, setting-appropriate studies are required to evaluate clinical outcomes.

Authors' Contributions

JT, BB, MO, JE, MCH, SG, and ET contributed to planning (study conception and protocol development). JE supervised the study and is sponsor of the project. JT and PDS conducted the study. PDS was the study physician. JT, BB, and SG contributed to the data analysis and interpretation. JT and BB wrote the manuscript. PDS, JE, MO, SG, and MCH revised the manuscript.

Conflicts of Interest

BB, MO, SG, ET, and MCH are or were employees, contractors, or equity holders in Ada Health GmbH. All should be considered to have an interest in Ada Health GmbH

References

1. Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW. Patients' use of the internet for medical information. J Gen Intern Med 2002 Mar;17(3):180-185 [FREE Full text] [doi: 10.1046/j.1525-1497.2002.10603.x] [Medline: 11929503]
2. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
3. Berner E. What can be done to increase the use of diagnostic decision support systems? Diagnosis (Berl) 2014 Jan 01;1(1):119-123. [doi: 10.1515/dx-2013-0014] [Medline: 29539976]
4. Mesko B, Győrffy Z. The rise of the empowered physician in the digital health era: viewpoint. J Med Internet Res 2019 Mar 26;21(3):e12490 [FREE Full text] [doi: 10.2196/12490] [Medline: 30912758]
5. Riches N, Panagioti M, Alam R, Cheraghi-Sohi S, Campbell S, Esmail A, et al. The effectiveness of electronic differential diagnoses (DDX) generators: a systematic review and meta-analysis. PLoS One 2016;11(3):e0148991 [FREE Full text] [doi: 10.1371/journal.pone.0148991] [Medline: 26954234]
6. Middleton B, Sittig DF, Wright A. Clinical decision support: a 25 year retrospective and a 25 year vision. Yearb Med Inform 2018 Mar 06;25(S 01):S103-S116. [doi: 10.15265/iys-2016-s034]
7. Densen P. Challenges and opportunities facing medical education. Trans Am Clin Climatol Assoc 2011;122:48-58 [FREE Full text] [Medline: 21686208]
8. Berner E, Lande T. Overview of clinical decision support systems. In: Clinical Decision Support Systems. Cham: Springer; 2016.
9. Graber MA, VanScoy D. How well does decision support software perform in the emergency department? Emerg Med J 2003 Sep;20(5):426-428 [FREE Full text] [doi: 10.1136/emj.20.5.426] [Medline: 12954680]
10. Balla J, Heneghan C, Goyder C, Thompson M. Identifying early warning signs for diagnostic errors in primary care: a qualitative study. BMJ Open 2012 Sep 13;2(5):e001539 [FREE Full text] [doi: 10.1136/bmjopen-2012-001539] [Medline: 22983786]
11. Ridderikhoff J, van Herk E. A diagnostic support system in general practice: is it feasible? Int J Med Informatics 1997 Jul;45(3):133-143. [doi: 10.1016/s1386-5056(97)00022-1]
12. Ronicke S, Hirsch MC, Türk E, Larionov K, Tientcheu D, Wagner AD. Can a decision support system accelerate rare disease diagnosis? Evaluating the potential impact of Ada DX in a retrospective study. Orphanet J Rare Dis 2019 Mar 21;14(1):69 [FREE Full text] [doi: 10.1186/s13023-019-1040-6] [Medline: 30898118]
13. Kelly A, Keijzers G, Klim S, Graham CA, Craig S, Kuan WS, AANZDEM Study Group. Asia, Australia and New Zealand Dyspnoea in Emergency Departments (AANZDEM) study: rationale, design and analysis. Emerg Med Australas 2015 Jun;27(3):187-191. [doi: 10.1111/1742-6723.12397] [Medline: 25940885]

14. Federal Act on Research involving Human Beings. Fedlex. URL: https://www.fedlex.admin.ch/eli/cc/2013/617/en [accessed 2022-02-09]

15. Ordinance on clinical trials with the exception of clinical trials of medical devices. Fedlex. URL: https://www.fedlex.admin.ch/eli/cc/2013/643/en [accessed 2022-02-09]

16. Nayak B, Hazra A. How to choose the right statistical test? Indian J Ophthalmol 2011;59(2):85-86 [FREE Full text] [doi: 10.4103/0301-4738.77005] [Medline: 21350275]

17. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc Series B (Methodological) 2018 Dec 05;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]

18. Armstrong RA. When to use the Bonferroni correction. Ophthalmic Physiol Opt 2014 Sep;34(5):502-508. [doi: 10.1111/opo.12131] [Medline: 24697967]

19. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]

20. Elkin PL, Liebow M, Bauer BA, Chaliki S, Wahner-Roedler D, Bundrick J, et al. The introduction of a diagnostic decision support system (DXplain™) into the workflow of a teaching hospital service can decrease the cost of service for diagnostically challenging Diagnostic Related Groups (DRGs). Int J Med Inform 2010 Nov;79(11):772-777 [FREE Full text] [doi: 10.1016/j.ijmedinf.2010.09.004] [Medline: 20951080]

21. Müller T, Jerrentrup A, Schäfer JR. [Computer-assisted diagnosis of rare diseases]. Internist (Berl) 2018 Apr;59(4):391-400. [doi: 10.1007/s00108-017-0218-z] [Medline: 28364276]

Abbreviations

DDSS: diagnostic decision support system
ED: emergency department

Original Paper

Effect of Information and Communication Technology–Based Self-management System DialBeticsLite on Treating Abdominal Obesity in the Specific Health Guidance in Japan: Randomized Controlled Trial

Masahiro Kondo[1,2], BHSc; Teru Okitsu[3], MD, PhD; Kayo Waki[4,5], MD, MPH, PhD; Toshimasa Yamauchi[5], MD, PhD; Masaomi Nangaku[6], MD, PhD; Kazuhiko Ohe[1], MD, PhD

[1]Department of Planning, Information and Management, University of Tokyo Hospital, Tokyo, Japan

[2]Graduate School of Interdisciplinary Information Studies, The University of Tokyo, Tokyo, Japan

[3]Institute of Industrial Science, The University of Tokyo, Tokyo, Japan

[4]Department of Biomedical Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[5]Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[6]Division of Nephrology and Endocrinology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

Corresponding Author:
Kayo Waki, MD, MPH, PhD
Department of Biomedical Informatics
Graduate School of Medicine
The University of Tokyo
7-3-1 Hongo
Bunkyo-ku
Tokyo, 113-8654
Japan
Phone: 81 3 5800 9129
Fax: 81 3 5800 9129
Email: kwaki-tky@m.u-tokyo.ac.jp

Abstract

Background: Mobile health (mHealth) interventions, a more cost-effective approach compared with traditional methods of delivering lifestyle coaching in person, have been shown to improve physical parameters and lifestyle behavior among overweight populations. In Japan, the Specific Health Checkups and Specific Health Guidance (SHG) started in 2008 to treat obesity and abdominal obesity. However, the effectiveness of SHG is limited owing to its in-person counseling. The effect of mHealth on SHG has yet to be demonstrated.

Objective: This study aims to determine whether a mobile self-management app (DialBeticsLite) could make the SHG more beneficial among patients with abdominal obesity to achieve a reduction in visceral fat area (VFA).

Methods: This study was an open-label, 2-arm, parallel-design randomized controlled trial. We recruited 122 people in September 2017 and randomly assigned them into either the intervention or control group. All participants attended an educational group session that delivered information regarding diet and exercise. In addition, participants in the intervention group were asked to use DialBeticsLite for 3 months. DialBeticsLite facilitated the daily recording of several physical parameters and lifestyle behavior and provided feedback to encourage an improvement in behavior. The primary outcome was the change in VFA from baseline to the 3-month follow-up. Secondary outcomes included changes in both physical and metabolic parameters from baseline to the 3-month follow-up. The Welch 2-tailed *t* test was conducted to analyze the effects of DialBeticsLite on both the primary and secondary outcomes.

Results: Of the 122 participants recruited, 75 (61.5%) were analyzed because 47 (38.5%) were excluded: 37 (30.3%) because of ineligibility and 10 (8.2%) because of withdrawal of consent. The mean age was 49.3 (SD 6.1) years in the intervention group (41/75, 55%) and 48.5 (SD 5.3) years in the control group (34/75, 45%), and all participants were men, although unintentionally. The baseline characteristics did not differ significantly between the intervention and control groups, except for VFA. The average change of VFA was −23.5 (SD 20.6) cm^2 in the intervention group and +1.9 (SD 16.2) cm^2 in the control group (*P*<.001).

Statistically significant differences were also found for the change of body weight, BMI, and waist circumference. These findings did not change after adjusting for VFA at the baseline. The intervention had no significant effect on any of the metabolic parameters. An exploratory analysis showed significant associations between the change in VFA and steps per day and between the change in VFA and calorie intake per day within the intervention group.

Conclusions: Our findings indicate that an mHealth intervention facilitating the daily monitoring of several physical parameters and lifestyle behavior can be highly effective in inducing visceral fat loss and weight loss among adults eligible for SHG.

Trial Registration: UMIN Clinical Trials Registry UMIN000042045; https://tinyurl.com/4vat3v53

Introduction

Background

Obesity is an increasingly concerning epidemic that is in need of urgent management [1], as its prevalence and the comorbidities and mortality associated with it have been, and continue to, rise rapidly worldwide [2-6]. It has also tremendously increased the global health burden and health care cost [1,5,7,8]. Abdominal obesity (AO), which is defined by the amount of visceral fat, is especially known as a risk factor for cardiovascular disease, diabetes, and fatty liver disease [9-12]. Lifestyle and behavioral interventions are considered effective for improving AO [13,14]. In recent years, nationwide lifestyle interventions have been implemented in several countries [15,16]. In Japan, the Specific Health Checkups and Specific Health Guidance started in 2008. All health insurers in Japan were required to offer health screening programs to all enrollees and their dependents aged between 40 and 74 years and to provide lifestyle counseling to participants who were not taking medication but who had AO or obesity and were at risk for metabolic syndrome. These programs have been shown to successfully lead to weight loss and improvements in several other physical parameters [16]. However, behavior changes are usually difficult to achieve in wide-scale clinical practice because of limited resources and professional support [17,18]. The Specific Health Guidance suffers from suboptimal participation, with only 23.2% of those asked to enroll in the guidance completing the course in 2019 [19]. There is the problem that there is great disparity of implementation rates between different insurers, as well as a likely variation of intervention intensity due to the relatively loose requirements for the content of support sessions.

These issues necessitate an effective self-management tool that can automate and standardize much of the counseling process [20]. In this regard, strategies using information and communication technology (ICT) could have great potential for the development of an effective and scalable lifestyle intervention. ICT-based interventions offer patients various benefits depending on the intervention design, including easier self-monitoring, access to health-related information, and personalized feedback, all contributing to optimizing the impact and convenience of the intervention [21,22]. They are also expected to minimize health care providers' workload and costs [21,23-25]. Previous ICT-based interventions have been shown to successfully reduce physical parameters such as body weight (BW), BMI, and waist circumference (WC) among overweight adults [21,22,26].

Objectives

In Japan, several companies and local governments started approaches to using ICT for the Specific Health Guidance [27,28]. However, to our knowledge, the effectiveness of ICT for the participants of the Specific Health Guidance has not been adequately evaluated in randomized controlled trials. Our primary aim is to investigate whether an app-based intervention was effective for people who are eligible for the Specific Health Guidance, the prevention program for lifestyle diseases in Japan. The authors of this study have developed an ICT-based self-management system, *DialBeticsLite*. This is a mobile app that allows the input of data on blood glucose, blood pressure (BP), BW, pedometer counts, diet, and physical exercise. It also provides users with fully automated evaluative feedback on diet modification and amount of physical exercise following the input of data each day. We provided some devices for people who use DialBeticsLite to measure their daily data. DialBetics, its former version, has been shown to successfully improve hemoglobin A_{1c} (HbA_{1c}) and fasting blood glucose levels in previous trials for patients with type 2 diabetes [29,30].

Methods

Study Design

This study was designed to be an open-label, 2-arm, parallel-design randomized controlled trial and was conducted at the University of Tokyo Hospital, Japan. The main objective is to evaluate the efficacy of treating AO using DialBeticsLite in a population who were eligible for the Specific Health Guidance.

Ethics Approval

The protocol and forms of informed consent were submitted to and approved by the Research Ethics Committee of The University of Tokyo Graduate School of Medicine and affiliated institutions (11696-[3]), and this study was carried out in compliance with the Declaration of Helsinki. This study was conducted among volunteer participants in a nonclinical setting, that is, a company, and was retrospectively registered at the University Hospital Medical Information Network Clinical Trials Registry (UMIN000042045).

XSL·FO

RenderX

Participants

The selection and recruitment of participants were conducted among the employees of a securities company located in Tokyo by advertising via email and the company intranet. This company is one of the top-ranked securities companies in Japan, and the average annual salary of employees is approximately JPY ¥10,000,000 (US $85,000). We recruited participants of the Specific Health Guidance who were also smartphone users and were willing to use the ICT system. Participants of the Specific Health Guidance were aged 40 to 75 years, who also satisfied both the following conditions: (1) WC ≥85 cm (men) or WC ≥90 cm (women) or BMI ≥25 kg/m^2 and (2) at least one of the three metabolic abnormalities from (1) hyperglycemia (fasting glucose level ≥100 mg/100 mL or HbA$_{1c}$ ≥5.6%), (2) hypertriglyceridemia (triglycerides ≥150 mg/dL) or low high-density lipoprotein (HDL)-cholesterol (HDL <40 mg/dL), and (3) high BP (systolic BP ≥130 mm Hg or diastolicBP ≥85 mm Hg). Patients who received any medication for the treatment of hypertension, dyslipidemia, or diabetes were excluded.

All participants who met the eligibility criteria had AO (WC ≥85 cm for men or ≥90 cm for women). The WC values used had been measured for screening AO at an annual routine health checkup.

All participants provided written informed consent before the trial commenced; they were informed of their right to withdraw from the study at any time and how the data collected from the study would be used. This included an explanation of how data would be accessible to the research team to be used for analysis and dissemination, following the conclusion of the trial. Any results obtained from analyzing these data were to be presented at major domestic and international scientific conferences and submitted for peer-reviewed journals of international repute and visibility.

Design of DialBeticsLite

The details of the DialBeticsLite system are shown in Figure 1. The app facilitated the daily recording of several physical parameters, in addition to tracking lifestyle behavior, that is, diet and exercise. Participants were asked to measure their blood glucose and BP levels twice a day at home, once after waking up in the morning and once before going to bed at night. They were also required to measure their BW in the morning. We also asked them to wear a pedometer for the entire day to measure the number of daily steps and approximate the energy expended by walking. These data were then transferred to the participant's smartphone by near-field communication or Bluetooth and sent instantly to the server, where the data were automatically evaluated. The evaluation of pedometer counts was carried out following the Japanese official physical activity guidelines for health promotion, which sets the target count >8000 steps per day [31]. Other data including blood glucose and BP levels were evaluated based on the Japan Diabetes Society guideline's target values [32]; desired values were set at (1) blood glucose concentrations <110 mg/dL before breakfast and <140 mg/dL at bedtime, and (2) BP with systolic BP <125 mm Hg and diastolic BP <75 mm Hg. Once DialBeticsLite determined whether each reading was satisfactory according to the guidelines, the evaluation outcomes were received by each participant's smartphone immediately. When DialBeticsLite detected critical values including (1) blood glucose concentrations >400 mg/dL or <70 mg/dL and (2) BP with systolic BP >220 mm Hg or diastolic BP >110 mm Hg, they were automatically reported to the research team, and if deemed necessary, attending physicians were informed to contact a participant with abnormal data as appropriate.

In addition, the participants could also input dietary information (type and quantity of food, accompanied by a photograph of the meal) and details of the type and duration of physical exercise, which had been completed each day separately from pedometer counts. The app evaluated these data on lifestyle similarly to blood glucose concentrations and BP, with specific advice on diet modification sent to the participant immediately after the input of data. The app provided feedback on whether the participants' macronutrient balance was appropriate, along with visualizations that indicated the nutritional balance of each meal. Specific guidance regarding dietary fiber and salt was also provided, in an effort to maintain intake within the recommendation set out by the Japan Diabetes Society guidelines.
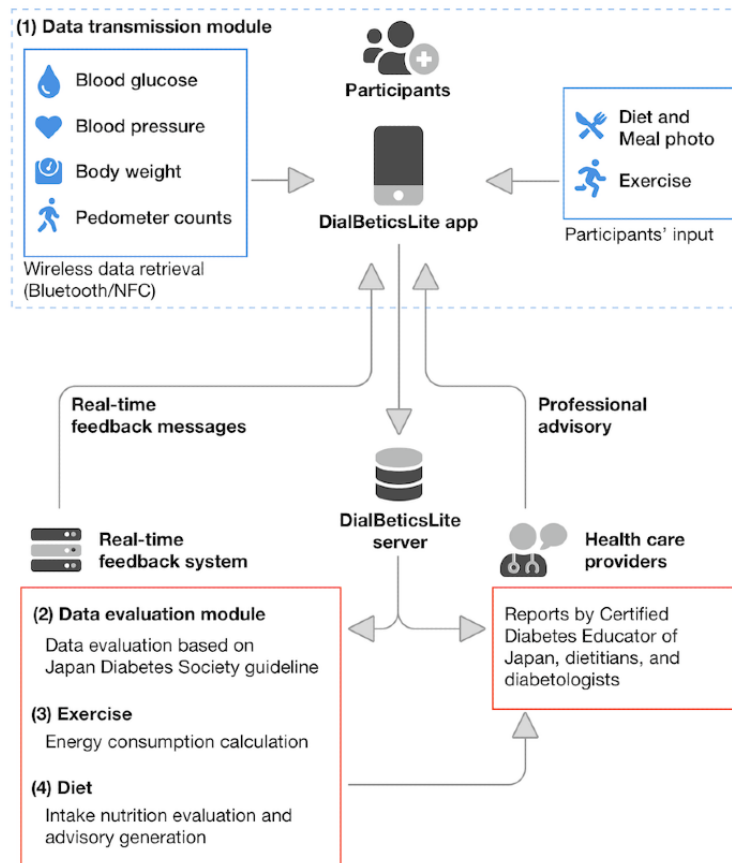
Intervention and Control

Blocked randomization within strata [33] was used to allocate participants equally to the intervention and control groups. We separated the participants into 2 BMI groups (BMI <25 kg/m^2 and BMI ≥25 kg/m^2), 2 visceral fat area (VFA) groups (VFA <100 cm^2 and VFA ≥100 cm^2), and 2 WC groups (WC <85 cm and WC ≥85 cm). This gave us 2×2×2=8 strata, and the block size in each stratum was 4. In other words, 4 participants in the same strata at a time were randomized to ensure that 2 patients were assigned to the treatment group and 2 patients were assigned to the control group.

All participants attended an educational group session at baseline, which was a 40-minute lecture at the University of Tokyo Hospital. This lecture was conducted by nurses, diabetologists, and dietitians who were all part of the research team and included information regarding diet and exercise.

The participants in the intervention group were asked to use DialBeticsLite for the next 3 months and trained on how to use the app by the research team after the lecture to ensure correct and informed use of the app (ie, how to take measurements and record them as intended, as well as how to interpret the readings). The intervention group was also provided with various other devices, including a Bluetooth-enabled BP monitor (HEM-7271T; Omron) and scale (HBF-255T; Omron), a near-filed communication–enabled glucometer (MS-FR201B; Terumo), and a pedometer (MT-KT02DZ; Terumo). These devices were all paired with a single smartphone (Galaxy Note3

SC-01F; Samsung) also provided as part of the intervention, which transmitted the readings to the DialbeticsLite server via a wireless network. For the full duration of the study, the research team provided technical assistance to the participants in the intervention group in addition to monitoring the content and frequency of daily records. The system automatically triggered an alert if participants failed to record data for >3 days, and these alerts were checked by the research team via the administrator screen. The nurse encouraged participants to restart measurements, either by email (after 1 week of inactivity) or by giving a call on the phone (after an additional week of inactivity). In cases where a participant failed to record any data (ie, not recording any of the physical parameters, food, or exercise) for 3 weeks, we deemed the participant as a dropout. As the system was solely designed for self-management and direct feedback using the input data, participants were asked to consult the physicians of their company if they had any health concerns unrelated to the intervention.

In contrast, the participants in the control group were not provided with the app or electronic devices, and they only participated in the assessment and lecture at baseline, in addition to the follow-up assessments at the end of 3 months.

Study Outcomes and Data Collection

The primary outcome of this study was the change in VFA from baseline to the 3-month follow-up. VFA was measured by differentiating visceral fat and abdominal subcutaneous fat using the current flow from 2 routes (DUALSCAN, HDS-2000; Fukuda Colin). HDS-2000 underestimates VFA compared with

computed tomography scan, but the correlation was very high (r=0.89) [34]. HDS-2000 can be a good alternative for evaluating VFA because of its simplicity and noninvasiveness. To prevent variation across raters in the measurement procedures, the VFA of all participants was measured by the same individual with sufficient experience. The person was unblinded for the group of intervention. Secondary outcomes were changes in both physical and metabolic parameters from baseline to the 3-month follow-up. Physical parameters included BW, WC, BMI, and BP. Metabolic parameters included cholesterol, triglyceride, fasting plasma glucose, and HbA$_{1c}$ levels, which were measured using blood tests. The physical parameters and metabolic parameters of all participants were measured at baseline and at the 3-month follow-up at the University of Tokyo Hospital. BP was measured after the participants took two deep breaths in the sitting position, whereas WC was measured at the umbilical level in the standing position. BMI was determined by calculating the ratio of BW (kg) to height squared (m^2). Blood tests were used to measure the concentration of fasting plasma glucose, HbA$_{1c}$, triglyceride, total cholesterol, low-density lipoprotein (LDL) cholesterol, and HDL cholesterol.

For the intervention group, the following variables were also assessed: total energy intake (kcal), steps taken per day, the use rate of the app, and the use rates of individual functions within the app, including blood glucose, BP, BW, diet, and physical exercise.

Statistical Analysis

The demographic characteristics and other parameters of the intervention and the control groups at baseline were presented as mean (SD) and compared using the Welch 2-tailed t test. To evaluate the effect of the DialbeticsLite, we used the Welch t test to compare the change in VFA (primary outcome) and changes in other parameters (secondary outcomes) between the 2 groups. We also conducted a linear regression analysis for both the primary and secondary outcomes, adjusted for VFA at baseline, as a post hoc test.

We also conducted several exploratory analyses to examine whether the improvement in AO was associated with lifestyle factors (pedometer counts and calorie intake) or the different use patterns of DialBeticsLite observed within the intervention group. We presented the numbers of days for which each parameter was recorded (ie, the number of days each function was used for) as the median (IQR). We calculated the Pearson correlation index and conducted tests for noncorrelation for each function to examine the relationship between the change in VFA and the use rate of each function. We plotted the trend of the average pedometer counts and calorie intake per day over the study period. We performed linear regression analysis to assess the effect of average pedometer counts and calorie intake per day on the change in VFA, adjusted for age and VFA at baseline. The retention rates of the pedometer count and diet functions were also plotted to help visualize engagement and retention over time.

A P value of <.05 was considered statistically significant, and all statistical analyses were performed using available-case analysis and using SAS (version 9.4; SAS Institute Inc).

Results

Participants

The study was approved in September 2017, and participants were recruited. The final 3-month follow-up was completed in February 2018. As shown in Figure 2, a total of 122 participants who provided written informed consent were randomly assigned to either the intervention or control group at randomization. This initially resulted in 50% (61/122) of participants for each group. Upon reassessment of eligibility, of the 61 participants, 20 (33%) were excluded from the intervention group: 18 (30%) owing to ineligibility and 2 (3%) owing to withdrawal of consent. Similarly, for the control group, of the 61 participants, 27 (44%) were excluded: 19 (31%) owing to ineligibility and 8 (13%) owing to withdrawal of consent. Furthermore, 2 participants dropped out over the course of the trial and were excluded from the intervention group.

There was a significant difference in VFA between the intervention and control groups at baseline but no difference in the other variables (Table 1). Although there was no intention of studying well-educated men exclusively, all the participants who participated in this trial were men university graduates.

Figure 2. Flow diagram of the study participants. VFA: visceral fat area; WC: waist circumference.
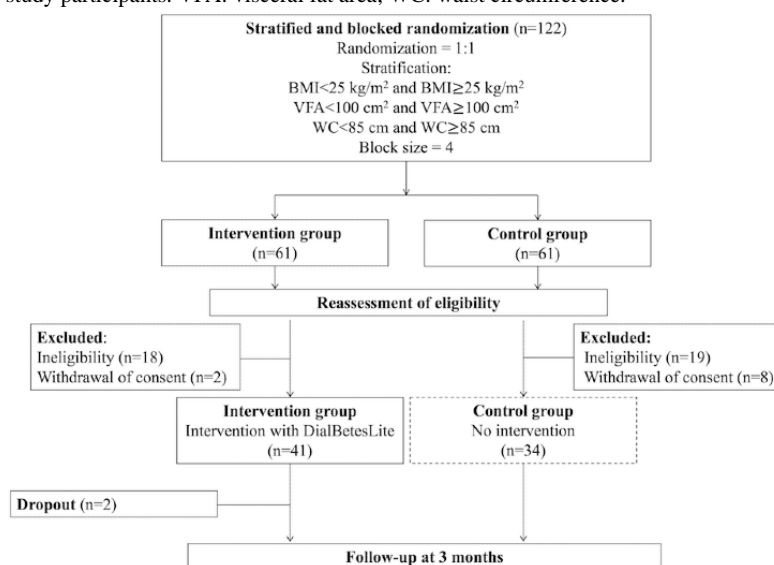
Table 1. Baseline characteristics of participants.

Variable	Intervention group (n=41)	Control group (n=34)	*P* value
Age (years), mean (SD)	49.3 (6.1)	48.5 (5.3)	.55
Sex, n (%)			
Men	41 (100)	34 (100)	__a
Women	0 (0)	0 (0)	—
Physical parameters, mean (SD)			
Visceral fat area (cm^2)	118.9 (32.3)	105.4 (20.8)	.03
Height (cm)	172.2 (5.0)	171.3 (6.5)	.49
Body weight (kg)	81.3 (9.9)	78.0 (6.8)	.09
BMI (kg/m^2)	27.4 (3.0)	26.6 (2.2)	.19
Waist circumference (cm)	96.1 (7.7)	95.0 (9.2)	.57
BP^b, mean (SD)			
Systolic BP (mm Hg)	130.7 (10.7)	128.2 (10.2)	.32
Diastolic BP (mm Hg)	86.3 (8.4)	86.9 (7.3)	.76
Lipid metabolism and glucose, mean (SD)			
Total cholesterol (mg/dL)	217.8 (30.4)	221.0 (37.3)	.69
LDL^c cholesterol (mg/dL)	137.7 (24.0)	141.1 (34.8)	.64
HDL^d cholesterol (mg/dL)	51.5 (10.0)	52.2 (11.2)	.76
Triglyceride (mg/dL)	142.9 (75.9)	138.3 (78.7)	.80
Fasting plasma glucose (mg/dL)	88.9 (12.1)	89.5 (7.7)	.80
HbA_{1c}^e (%)	5.5 (0.3)	5.5 (0.3)	.85

[a]Comparisons of sex were not conducted because all participants were males.

[b]BP: blood pressure.

[c]LDL: low-density lipoprotein.

[d]HDL: high-density lipoprotein.

[e]HbA_{1c}: hemoglobin A_{1c}.

Changes in VFA and Physical and Metabolic Parameters

At the 3-month follow-up, although the control group observed an increase in VFA (mean +1.9, SD 16.2 cm^2), participants in the intervention group lost considerable amounts of VFA (mean −23.5, SD 20.6 cm^2). Consequently, those who underwent the intervention had statistically significant reductions in the primary outcome compared with those in the control group ($P<.001$). This finding did not change after adjusting for VFA at baseline between the 2 groups as a post hoc test (Table 2).

In addition, the intervention group displayed a statistically significant improvement in BW (mean −3.0, SD 2.8 kg vs mean+1.1, SD 1.6 kg; $P<.001$), BMI (mean −1.0, SD 1.0 kg/m^2 vs mean+0.4, SD 1.6 kg/m^2; $P<.001$), and WC (mean −4.8, SD 3.8 cm vs mean −1.6, SD 7.7 cm; $P=.04$) compared with those in the control group. For all of the remaining physical and metabolic parameters that were assessed, including BP, cholesterol, triglycerides, fasting plasma glucose, and HbA$_{1c}$, no differences between the 2 groups were observed (Table 2).

Further analysis revealed that the reduction in VFA in the intervention group was significantly associated with both the number of steps per day ($P<.001$) and calorie intake per day ($P<.001$), after results were adjusted to take age and the VFA at baseline into consideration (Table 3).

Table 2. Change in parameters.

Variable	Intervention group (n=39), mean (SD)	Control group (n=34), mean (SD)	P value	Adjusted P value[a]
Physical parameters				
Visceral fat area (cm^2)	−23.5 (20.6)	1.9 (16.2)	<.001	<.001
Body weight (kg)	−3.0 (2.8)	1.1 (1.6)	<.001	<.001
BMI (kg/m^2)	−1.0 (1.0)	0.4 (0.6)	<.001	<.001
Waist circumference (cm)	−4.8 (3.8)	−1.6 (7.7)	.04	.02
BP[b]				
Systolic BP (mm Hg)	−3.9 (10.1)	−1.4 (9.6)	.29	.28
Diastolic BP (mm Hg)	0.1 (9.0)	0.9 (8.5)	.68	.61
Lipid metabolism and glucose				
Total cholesterol (mg/dL)	−4.1 (20.1)	2.0 (21.7)	.22	.33
LDL[c] cholesterol (mg/dL)	−3.3 (16.1)	−1.6 (21.6)	.72	.53
HDL[d] cholesterol (mg/dL)	4.5 (7.9)	1.4 (5.7)	.06	.07
Triglyceride (mg/dL)	−26.4 (50.5)	11.5 (88.3)	.03	.11
Fasting plasma glucose[e] (mg/dL)	1.1 (6.7)	2.7 (7.3)	.35	.55
HbA$_{1c}$[f] (%)	0.0 (0.1)	0.0 (0.1)	.72	.97

[a]Adjusted for visceral fat area at baseline.

[b]BP: blood pressure.

[c]LDL: low-density lipoprotein.

[d]HDL: high-density lipoprotein.

[e]Because 1 participant was absent in the control group, the number of participants analyzed was 39 in the intervention group and 33 in the control group.

[f]HbA$_{1c}$: hemoglobin A$_{1c}$.

Table 3. Linear regression for the change in visceral fat area in the intervention group.

| Variable	Estimate	95% CI	P value
Number of steps per day (every 1000 steps)	−4.76	−6.59 to −2.92	<.001
Calorie intake per day (every 100 kcal)	2.29	1.00 to 3.58	<.001
Age (years)	−0.33	−1.14 to 0.48	.42
Visceral fat area at baseline (cm^2)	−0.11	−0.25 to 0.03	.13

XSL•FO
RenderX

Changes of Number of Steps and Calorie Intake During the Study Period

In the intervention group, participants walked an average of nearly 8000 (SD 5139) steps (Figure 3). Similarly, the average calorie intake was nearly 2000 (SD 703) kcal/day (Figure 3). Both parameters remained almost constant during the study period.

Figure 3. Variables of lifestyle behavior recorded during the study period. (A) Number of steps. Solid line: daily mean pedometer counts; dotted lines: mean (SD). (B) Calorie intake. Solid line: daily mean calorie intake; dotted lines: mean (SD).
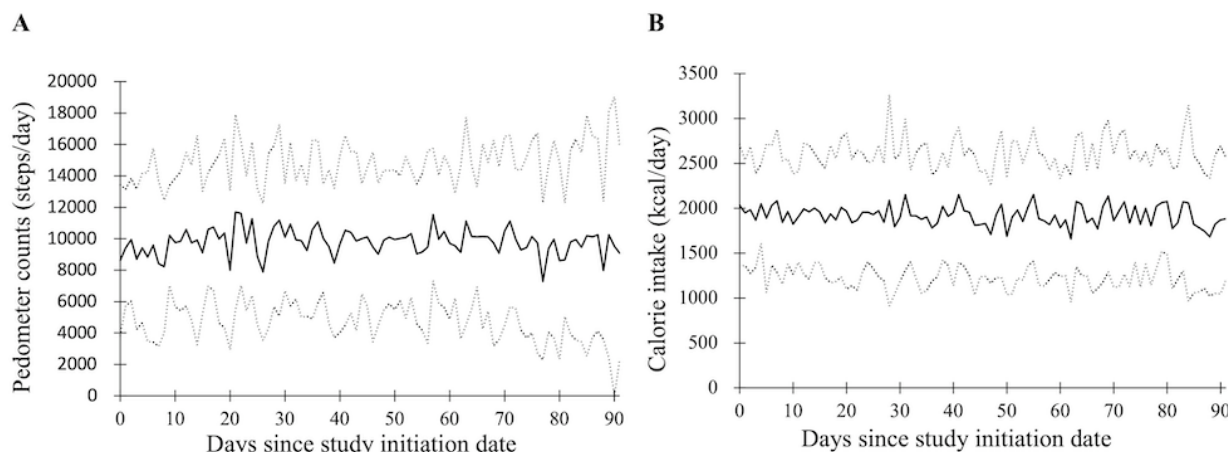
Use of DialBeticsLite in the Intervention Group

By analyzing the data recorded by DialBeticsLite, we found the median use rate of the app to be 100% (IQR 98%-100%) within the intervention group (Table 4). When examining individual functions, differences in use rates were observed. The use rates of the self-monitoring functions for BW (median 93%, IQR 73%-98%), pedometer counts (median 85%, IQR 74%-93%), BP before breakfast (median 88%, IQR 61%-95%), breakfast (median 98%, IQR 78%-100%), lunch (median 97%, IQR 62%-99%), and dinner (median 95%, IQR 75%-99%) were high, whereas those regarding exercise input (median 1%, IQR 0%-14%), BP at bedtime (median 61%, IQR 29%-86%), blood glucose (median 0%, IQR 0%-0%), and snacks (median 20%, IQR 10%-36%) were low. We examined the use of functions that recorded lifestyle and behavior, specifically the pedometer count function and the diet function (including breakfast, lunch,

dinner, and snacks). We found that both the retention rate of the pedometer counts function and diet function decreased over time during the study, with the former declining from 90% to 50% and the latter declining from 90% to 70% (Figure 4). We analyzed the recorded data by producing a Pearson correlation matrix to determine whether any of the individual function's use rates showed a correlation with the reduction of VFA (Multimedia Appendix 1). We found that the BW ($r=-0.27$, $P=.10$), pedometer counts ($r=-0.25$, $P=.13$), BP before breakfast ($r=-0.27$, $P=.10$), and snacks ($r=-0.27$, $P=.10$) functions showed a positive correlation with the reduction of VFA, although this was not statistically significant. In contrast, a significant correlation was observed between the use rates of the functions. Correlation coefficients of over 0.9 were obtained between the BW and BP before breakfast functions, as well as between the individual diet functions, including breakfast and lunch, breakfast and dinner, and lunch and dinner.

Table 4. Number of days with recordings and use rate for each parameter in the intervention group.

Variable	Days with recordings (n=92), median (IQR)	Use rate (%), median (IQR)
Total[a]	92 (90-92)	100 (98-100)
Body weight	86 (67-90)	93 (73-98)
Pedometer counts	78 (68-86)	85 (74-93)
Exercise input	1 (0-13)	1 (0-14)
Blood pressure		
Before breakfast	81 (56-87)	88 (61-95)
At bedtime	56 (27-79)	61 (29-86)
Blood glucose		
Before breakfast	0 (0-0)	0 (0-0)
At bedtime	0 (0-0)	0 (0-0)
Calorie intake		
Breakfast	90 (72-92)	98 (78-100)
Lunch	89 (57-91)	97 (62-99)
Dinner	87 (69-91)	95 (75-99)
Snacks	18 (9-33)	20 (10-36)

[a]The median number of days with at least one variable recorded.

Figure 4. Retention rates of the functions regarding lifestyle behavior during study period. (A) Pedometer counts and (B) calorie intake.
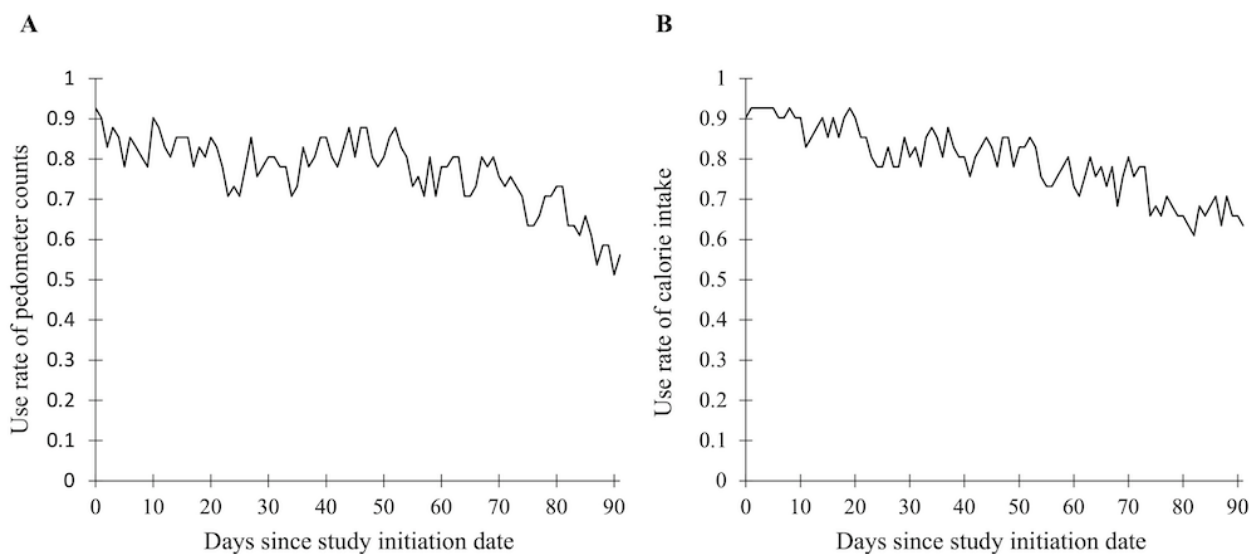
Discussion

Principal Findings

In this randomized clinical trial, we examined the effect of using DialBeticsLite for approximately 3 months in addition to traditional counseling for the participants of the Specific Health Guidance with AO. Statistically significant reductions were observed in VFA, which was the primary end point, and some other parameters (BW, BMI, and WC). These results suggest that the addition of mHealth interventions could make the Specific Health Guidance more beneficial. According to an exploratory analysis within the intervention group, the mean number of steps and calorie intake per day during the study period were associated with a decrease in VFA.

We observed a significant reduction for VFA, BW, BMI, and WC after the intervention, which is consistent with other studies that have shown that mHealth interventions can reduce BW, BMI, and WC [21,22,26]. DialBeticsLite encourages self-monitoring through daily records and provides personalized advice. Frequent self-weighing and recording of daily weight patterns by patients with obesity has been reported to be effective in weight loss programs [35]. Our findings also reinforce the effectiveness of self-monitoring for individuals with AO. Owing to the nature of the intervention, its impact relies on engagement [36]. Previous studies have indicated so, reporting a significant association between weight loss and the use rate of self-monitoring functions or the use rate of an app in general [36,37]. The same was observed for behavior, with

participants with higher app use exhibiting increased physical activity and decreased caloric intake [37]. In this study, although strong correlations were not observed between the loss of VFA and the use of self-monitoring functions, they do not necessarily contradict previous studies. Engagement in the intervention group was higher than that in other studies involving behavioral self-monitoring with mHealth, as the median number of days of full app use was 92 out of 92 days [22,38]. Similarly, a median of over 80 days for both BW and diet self-monitoring functions (excluding snacks) was observed, which was also higher than that in past trials [29]. The higher engagements may have resulted in difference of reduction of VFA between the intervention and control groups.

Another important finding, but perhaps least surprising, was that VFA reduction was associated with the number of steps taken and calorie intake in the intervention group. Increasing the number of steps and decreasing the number of calories are natural ways to improve AO. In the intervention group, participants walked an average of nearly 8000 steps (Figure 3), which seems to be sufficient, but an increase in the number of steps may still be effective. Similarly, further reduction of calorie intake from 2000 kcal/day (Figure 3) may further improve AO. Although some randomized controlled trials have shown that mHealth interventions are effective [21,22,26], few studies have revealed the cause of the improvement. Our findings may support the idea that we should focus not only on increasing engagement but also on causing behavior change.

In contrast to the significant reduction for VFA, BW, BMI, and WC after the intervention, all remaining physical and metabolic parameters, including BP, cholesterol, triglycerides, fasting plasma glucose, and HbA_{1c}, showed no differences between the 2 groups. However, because these parameters were well controlled and almost within the normal range at baseline, improvement may have been difficult regardless of the intervention.

Although mHealth allows for easier initiation and recruitment, it is also more prone to disengagement [39]. Over the course of this study, the retention rates declined within the intervention group (Figure 4). This is also seen in other studies, and there is a need to explore potential methods to increase engagement and retention to maximize the benefits and efficacy of mHealth as a replacement for in-person support [36]. A study examining the reasons for dropout identified that daily recording could become overly repetitive and burdensome, with too much data to track [38], whereas another study targeting patients who had been prescribed medication discovered that the time and effort required were the predominant barriers to using personal health record systems [40]. In contrast, the frequent use of self-monitoring functions has been shown to increase retention [41], and low engagement has been identified as a predictor for participants abandoning the app [42]. Therefore, retention and adherence could potentially be improved by identifying specific functions of the app with higher use rates and associations with positive outcomes while eliminating other features to simplify the app. For instance, in our study, the recording function for blood glucose was rarely used and may have been useless. In addition, although not statistically significant, the BW,

pedometer counts, BP before breakfast, and snacks functions showed a positive correlation with the reduction of VFA. The high measurement rates of this study population might have resulted in an insignificant correlation; placing further emphasis on and developing these functions might further improve the effect of the intervention on AO. Furthermore, the use rates of some functions were strongly associated with those of other functions. For example, those who used the BW function were more likely to have been using the BP before breakfast function, and the same was seen between the individual diet functions, including breakfast and lunch, breakfast and dinner, and lunch and dinner. To optimize the app design, further detailed investigations into the effect of each feature are required.

Limitations

Although the results of this study are promising, they should be interpreted with some caution, as there are several limitations to this study. First, past studies have revealed that for studies examining the effect of lifestyle interventions, those with sample sizes of >90 and intervention durations of >8 weeks tend to have more successful outcomes [36]. Although, at randomization, our study included 122 individuals, only 73 (59.8%) were included in the final analysis of results, and this relatively small sample included only Japanese men aged 40 to 75 years. Many patients were found to not meet the eligibility criteria after allocation because of our lack of confirmation; assessment was determined only by the information obtained before allocation. Thus, the transportability of the results to other populations, including those with women and different ethnic groups, is limited. As there has been evidence from a recent study to support a difference in retention and willingness to use mHealth interventions between men and women, with the latter being more reluctant to use apps intended to treat diabetes [43], there is a need to assess whether DialBeticsLite also has an uneven adherence between the 2 sexes.

Second, the participants were observed for a relatively short period of 3 months, and although this is longer than 8 weeks, the long-term impact of the intervention remains to be tested. This is especially important considering that many participants in lifestyle interventions struggle to maintain the significant physical and behavior changes achieved at the end of the intervention [44,45]. There has been evidence of the effect of mHealth being relatively short-lived, with participants regaining any weight they had lost during the intervention period [46].

Third, despite using stratified randomization, VFA differed between the 2 groups at baseline. As we categorized the VFA and allocated participants by stratified randomization, this phenomenon has some probability of occurrence. However, a sensitivity analysis to examine the influence of this difference was conducted during the analysis of primary and secondary outcomes, and as the results were relatively unchanged, we concluded that the difference in VFA at baseline had little influence on the primary and secondary outcomes.

Fourth, the provision of some medical devices to the intervention group may have influenced the results obtained by improving the participants' motivation. Therefore, it may be more appropriate to interpret the results of this study as indicative of

the combined effect of the app and the distribution of medical devices, rather than the effect of the app alone.

Finally, as daily steps and caloric intake were not recorded in the control group, we cannot be certain of the extent to which these 2 factors were affected by the intervention. Although physical parameters improved significantly, we were unable to observe a significant reduction in total energy intake or an increase in physical activity during the intervention. At first glance, this may suggest that the intervention did not contribute to improving lifestyle and behavior; however, considering a previous study that showed that mHealth intervention prevented the reduction of daily steps, this may not be the case [47]. Although this previous study did not produce statistically significant results, likely owing to its small sample size, it suggests that an increase in daily steps is not necessary for an mHealth intervention to be meaningful. Allowing participants to maintain a high level of physical exercise in itself is an achievement and may explain why our study saw a significant decrease in BW, BMI, and VFA despite seeing no significant change for daily steps and total caloric intake.

Conclusions

Our study suggested that DialBeticsLite, a mobile app designed to assist self-management, was feasible among Japanese adults who were eligible for the Specific Health Guidance. The mHealth intervention resulted in a statistically significant reduction in VFA, BW, BMI, and WC over the course of 3 months. Compared with traditional methods, ICT systems offer greater scalability and convenience and could be more cost-effective. Thus, these findings are promising and show that mHealth interventions have great potential for treating patients with AO.

Conflicts of Interest

KW was a member of the Department of Ubiquitous Health Informatics when the study was conducted and developed the information and communication technology–based self-management system, DialBeticsLite.

Multimedia Appendix 1
Pearson correlation matrix for the change in visceral fat area and number of days with recordings for each parameter in the intervention group.
[PDF File (Adobe PDF File), 450 KB - formative_v6i3e33852_app1.pdf]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V1.6.1).
[PDF File (Adobe PDF File), 726 KB - formative_v6i3e33852_app2.pdf]

References

1.   Dai H, Alsalhe TA, Chalghaf N, Riccò M, Bragazzi NL, Wu J. The global burden of disease attributable to high body mass index in 195 countries and territories, 1990-2017: an analysis of the Global Burden of Disease Study. PLoS Med 2020 Jul;17(7):e1003198 [FREE Full text] [doi: 10.1371/journal.pmed.1003198] [Medline: 32722671]
2.   Abdelaal M, le Roux CW, Docherty NG. Morbidity and mortality associated with obesity. Ann Transl Med 2017 Apr;5(7):161 [FREE Full text] [doi: 10.21037/atm.2017.03.107] [Medline: 28480197]
3.   Çelmeli G, Çürek Y, Arslan Gülten Z, Yardımsever M, Koyun M, Akçurin S, Bircan. Remarkable increase in the prevalence of overweight and obesity among school age children in Antalya, Turkey, between 2003 and 2015. J Clin Res Pediatr Endocrinol 2019 Feb 20;11(1):76-81 [FREE Full text] [doi: 10.4274/jcrpe.galenos.2018.2018.0108] [Medline: 30251957]
4.   Agha M, Agha R. The rising prevalence of obesity: part A: impact on public health. Int J Surg Oncol (N Y) 2017 Aug;2(7):e17 [FREE Full text] [doi: 10.1097/IJ9.0000000000000017] [Medline: 29177227]
5.   Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999-2004. J Am Med Assoc 2006 Apr 05;295(13):1549-1555. [doi: 10.1001/jama.295.13.1549] [Medline: 16595758]
6.   Yang L, Colditz GA. Prevalence of overweight and obesity in the United States, 2007-2012. JAMA Intern Med 2015 Aug;175(8):1412-1413. [doi: 10.1001/jamainternmed.2015.2405] [Medline: 26098405]
7.   GBD 2015 Obesity Collaborators, Afshin A, Forouzanfar MH, Reitsma MB, Sur P, Estep K, et al. Health effects of overweight and obesity in 195 countries over 25 years. N Engl J Med 2017 Dec 06;377(1):13-27 [FREE Full text] [doi: 10.1056/NEJMoa1614362] [Medline: 28604169]

XSL·FO
RenderX

8. Dee A, Kearns K, O'Neill C, Sharp L, Staines A, O'Dwyer V, et al. The direct and indirect costs of both overweight and obesity: a systematic review. BMC Res Notes 2014 Apr 16;7:242 [FREE Full text] [doi: 10.1186/1756-0500-7-242] [Medline: 24739239]

9. Fang H, Berg E, Cheng X, Shen W. How to best assess abdominal obesity. Curr Opin Clin Nutr Metab Care 2018 Sep;21(5):360-365 [FREE Full text] [doi: 10.1097/MCO.0000000000000485] [Medline: 29916924]

10. Després JP, Lemieux I, Prud'homme D. Treatment of obesity: need to focus on high risk abdominally obese patients. Br Med J 2001 Mar 24;322(7288):716-720 [FREE Full text] [doi: 10.1136/bmj.322.7288.716] [Medline: 11264213]

11. Cerhan JR, Moore SC, Jacobs EJ, Kitahara CM, Rosenberg PS, Adami H, et al. A pooled analysis of waist circumference and mortality in 650,000 adults. Mayo Clin Proc 2014 Mar;89(3):335-345 [FREE Full text] [doi: 10.1016/j.mayocp.2013.11.011] [Medline: 24582192]

12. Sahakyan KR, Somers VK, Rodriguez-Escudero JP, Hodge DO, Carter RE, Sochor O, et al. Normal-weight central obesity: implications for total and cardiovascular mortality. Ann Intern Med 2015 Dec 01;163(11):827-835 [FREE Full text] [doi: 10.7326/M14-2525] [Medline: 26551006]

13. Gremeaux V, Drigny J, Nigam A, Juneau M, Guilbeault V, Latour E, et al. Long-term lifestyle intervention with optimized high-intensity interval training improves body composition, cardiometabolic risk, and exercise parameters in patients with abdominal obesity. Am J Phys Med Rehabil 2012 Nov;91(11):941-950. [doi: 10.1097/PHM.0b013e3182643ce0] [Medline: 22854902]

14. Ilanne-Parikka P, Eriksson JG, Lindström J, Peltonen M, Aunola S, Hämäläinen H, Finnish Diabetes Prevention Study Group. Effect of lifestyle intervention on the occurrence of metabolic syndrome and its components in the Finnish Diabetes Prevention Study. Diabetes Care 2008 Apr;31(4):805-807. [doi: 10.2337/dc07-1117] [Medline: 18184907]

15. Cannon MJ, Masalovich S, Ng BP, Soler RE, Jabrah R, Ely EK, et al. Retention among participants in the National Diabetes Prevention Program Lifestyle Change Program, 2012-2017. Diabetes Care 2020 Sep;43(9):2042-2049. [doi: 10.2337/dc19-2366] [Medline: 32616617]

16. Tsushita K, Hosler A, Miura K, Ito Y, Fukuda T, Kitamura A, et al. Rationale and descriptive analysis of specific health guidance: the Nationwide Lifestyle Intervention Program targeting metabolic syndrome in Japan. J Atheroscler Thromb 2018 Apr 01;25(4):308-322 [FREE Full text] [doi: 10.5551/jat.42010] [Medline: 29238010]

17. Caterson ID, Alfadda AA, Auerbach P, Coutinho W, Cuevas A, Dicker D, et al. Gaps to bridge: misalignment between perception, reality and actions in obesity. Diabetes Obes Metab 2019 Aug;21(8):1914-1924 [FREE Full text] [doi: 10.1111/dom.13752] [Medline: 31032548]

18. Dicker D, Alfadda AA, Coutinho W, Cuevas A, Halford JC, Hughes CA, et al. Patient motivation to lose weight: importance of healthcare professional support, goals and self-efficacy. Eur J Intern Med 2021 Sep;91:10-16 [FREE Full text] [doi: 10.1016/j.ejim.2021.01.019] [Medline: 33558163]

19. The state of implementation of specific health checkups and specific health guidance in FY 2019 (Japanese). Ministry of Health, Labor and Welfare. URL: https://www.mhlw.go.jp/content/12400000/000755573.pdf [accessed 2021-09-30]

20. Steinberg DM, Tate DF, Bennett GG, Ennett S, Samuel-Hodge C, Ward DS. The efficacy of a daily self-weighing weight loss intervention using smart scales and e-mail. Obesity (Silver Spring) 2013 Sep;21(9):1789-1797 [FREE Full text] [doi: 10.1002/oby.20396] [Medline: 23512320]

21. Thomas JG, Leahey TM, Wing RR. An automated internet behavioral weight-loss program by physician referral: a randomized controlled trial. Diabetes Care 2015 Jan;38(1):9-15. [doi: 10.2337/dc14-1474] [Medline: 25404659]

22. Fukuoka Y, Gay CL, Joiner KL, Vittinghoff E. A novel diabetes prevention intervention using a mobile app: a randomized controlled trial with overweight adults at risk. Am J Prev Med 2015 Aug;49(2):223-237 [FREE Full text] [doi: 10.1016/j.amepre.2015.01.003] [Medline: 26033349]

23. Castelnuovo G, Manzoni GM, Pietrabissa G, Corti S, Giusti EM, Molinari E, et al. Obesity and outpatient rehabilitation using mobile technologies: the potential mHealth approach. Front Psychol 2014;5:559 [FREE Full text] [doi: 10.3389/fpsyg.2014.00559] [Medline: 24959157]

24. Ryu S. Book review: mHealth: new horizons for health through mobile technologies: based on the findings of the second global survey on eHealth (Global Observatory for eHealth Series, Volume 3). Healthc Inform Res 2012;18(3):231. [doi: 10.4258/hir.2012.18.3.231]

25. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]

26. Lin P, Wang Y, Levine E, Askew S, Lin S, Chang C, et al. A text messaging-assisted randomized lifestyle weight loss clinical trial among overweight adults in Beijing. Obesity (Silver Spring) 2014 May;22(5):29-37. [doi: 10.1002/oby.20686] [Medline: 24375969]

27. Fukuoka 100 Project. URL: https://100.city.fukuoka.lg.jp/ [accessed 2022-02-10]

28. Remote-contained 'Specific Health Guidance Service' for health insurance companies, etc., exceeds 100 installations and 10,000 users. SMS Co., Ltd. URL: https://www.bm-sms.co.jp/news-press/prs_20200709_tokuteihokenshidou/ [accessed 2022-02-10]

29. Waki K, Fujita H, Uchimura Y, Omae K, Aramaki E, Kato S, et al. DialBetics: a novel smartphone-based self-management support system for type 2 diabetes patients. J Diabetes Sci Technol 2014 Mar 13;8(2):209-215 [FREE Full text] [doi: 10.1177/1932296814526495] [Medline: 24876569]

30. Waki K, Aizawa K, Kato S, Fujita H, Lee H, Kobayashi H, et al. DialBetics with a multimedia food recording tool, FoodLog: smartphone-based self-management for type 2 diabetes. J Diabetes Sci Technol 2015 May;9(3):534-540 [FREE Full text] [doi: 10.1177/1932296815579690] [Medline: 25883164]

31. Japanese official physical guidelines for health promotion. Ministry of Health, Labor and Welfare. 2014. URL: https://www.nibiohn.go.jp/eiken/info/pdf/active2013-e.pdf [accessed 2021-09-30]

32. Haneda M, Noda M, Origasa H, Noto H, Yabe D, Fujita Y, et al. Japanese clinical practice guideline for diabetes 2016. Diabetol Int 2018 Feb;9(1):1-45 [FREE Full text] [doi: 10.1007/s13340-018-0345-3] [Medline: 30603347]

33. Altman DG, Bland JM. How to randomise. Br Med J 1999 Sep 11;319(7211):703-704 [FREE Full text] [doi: 10.1136/bmj.319.7211.703] [Medline: 10480833]

34. Park KS, Lee D, Lee J, Kim YJ, Jung KY, Kim KM, et al. Comparison between two methods of bioelectrical impedance analyses for accuracy in measuring abdominal visceral fat area. J Diabetes Complicat 2016 Mar;30(2):343-349. [doi: 10.1016/j.jdiacomp.2015.10.014] [Medline: 26620129]

35. LaRose JG, Lanoye A, Tate DF, Wing RR. Frequency of self-weighing and weight loss outcomes within a brief lifestyle intervention targeting emerging adults. Obes Sci Pract 2016 Mar;2(1):88-92 [FREE Full text] [doi: 10.1002/osp4.24] [Medline: 27668087]

36. Schoeppe S, Alley S, Van LW, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]

37. Turner-McGrievy GM, Beets MW, Moore JB, Kaczynski AT, Barr-Anderson DJ, Tate DF. Comparison of traditional versus mobile app self-monitoring of physical activity and dietary intake among overweight adults participating in an mHealth weight loss program. J Am Med Inform Assoc 2013 May 1;20(3):513-518 [FREE Full text] [doi: 10.1136/amiajnl-2012-001510] [Medline: 23429637]

38. Griauzde D, Kullgren JT, Liestenfeltz B, Ansari T, Johnson EH, Fedewa A, et al. A mobile phone-based program to promote healthy behaviors among adults with prediabetes who declined participation in free diabetes prevention programs: mixed-methods pilot randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 09;7(1):e11267 [FREE Full text] [doi: 10.2196/11267] [Medline: 30626566]

39. Grock S, Ku J, Kim J, Moin T. A review of technology-assisted interventions for diabetes prevention. Curr Diab Rep 2017 Sep 23;17(11):107. [doi: 10.1007/s11892-017-0948-2] [Medline: 28942537]

40. Hirano R, Yamaguchi S, Waki K, Kimura Y, Chin K, Nannya Y, et al. Willingness of patients prescribed medications for lifestyle-related diseases to use personal health records: questionnaire study. J Med Internet Res 2020 May 28;22(5):e13866 [FREE Full text] [doi: 10.2196/13866] [Medline: 32463368]

41. Lee K, Kwon H, Lee B, Lee G, Lee JH, Park YR, et al. Effect of self-monitoring on long-term patient engagement with mobile health applications. PLoS One 2018;13(7):e0201166 [FREE Full text] [doi: 10.1371/journal.pone.0201166] [Medline: 30048546]

42. Scherer EA, Ben-Zeev D, Li Z, Kane JM. Analyzing mHealth engagement: joint models for intensively collected user engagement data. JMIR Mhealth Uhealth 2017 Jan 12;5(1):e1 [FREE Full text] [doi: 10.2196/mhealth.6474] [Medline: 28082257]

43. Yamaguchi S, Waki K, Nannya Y, Nangaku M, Kadowaki T, Ohe K. Usage patterns of GlucoNote, a self-management smartphone app, based on ResearchKit for patients with type 2 diabetes and prediabetes. JMIR Mhealth Uhealth 2019 Apr 24;7(4):e13204 [FREE Full text] [doi: 10.2196/13204] [Medline: 31017586]

44. Lindahl B, Nilssön TK, Borch-Johnsen K, Røder ME, Söderberg S, Widman L, et al. A randomized lifestyle intervention with 5-year follow-up in subjects with impaired glucose tolerance: pronounced short-term impact but long-term adherence problems. Scand J Public Health 2009 Jun;37(4):434-442. [doi: 10.1177/1403494808101373] [Medline: 19181821]

45. Stelmach-Mardas M, Mardas M, Walkowiak J, Boeing H. Long-term weight status in regainers after weight loss by lifestyle intervention: status and challenges. Proc Nutr Soc 2014 Nov;73(4):509-518. [doi: 10.1017/S0029665114000718] [Medline: 25192545]

46. Pinto S, Waki K, Chiovato L, De Cata P, Dagliati A, Tibollo V, et al. Smartphone-based self-management of non-insulin-dependent diabetes: a Japanese system at use by an Italian patients' cohort. J Diabetes Sci Technol 2018 Jul;12(4):903-904 [FREE Full text] [doi: 10.1177/1932296818763884] [Medline: 29633857]

47. Choi J, Lee JH, Vittinghoff E, Fukuoka Y. mHealth physical activity intervention: a randomized pilot study in physically inactive pregnant women. Matern Child Health J 2016 May;20(5):1091-1101. [doi: 10.1007/s10995-015-1895-7] [Medline: 26649879]

Abbreviations

AO: abdominal obesity

BP: blood pressure
BW: body weight
HbA$_{1c}$: hemoglobin A$_{1c}$
ICT: information and communication technology
mHealth: mobile health
VFA: visceral fat area
WC: waist circumference
HDL: high-density lipoprotein
LDL: low-density lipoprotein

XSL•FO
RenderX

<u>Original Paper</u>

Perspectives From French and Filipino Parents on the Adaptation of Child Health Knowledge Translation Tools: Qualitative Exploration

Sarah A Elliott[1,2], PhD; Kelsey S Wright[1], MSc; Shannon D Scott[3], PhD; Lisa Hartling[1,2], PhD

[1]Alberta Research Centre for Health Evidence, Department of Pediatrics, University of Alberta, Edmonton, AB, Canada

[2]Cochrane Child Health, Department of Pediatrics, University of Alberta, Edmonton, AB, Canada

[3]Faculty of Nursing, University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Lisa Hartling, PhD
Alberta Research Centre for Health Evidence
Department of Pediatrics
University of Alberta
87 Avenue NW
Edmonton, AB, T6G 1C9
Canada
Phone: 1 780 492 6248
Email: hartling@ualberta.ca

Abstract

Background: A number of evidence-based knowledge translation (KT) tools for parents of children with acute health conditions have been developed. These tools were created and tested with parental input and disseminated to groups proficient in English. Therefore, it is unclear whether they are useful for populations that are more diverse. To enhance the reach of our current and future KT tools, language translation and cultural adaptations may promote relevance for previously underserved knowledge users.

Objective: This study aims to explore and understand considerations for the cultural and linguistic adaptation of a KT tool in French and Filipino communities.

Methods: A KT tool (whiteboard animation video) describing the signs and symptoms of croup was originally developed in English to provide parents with evidence-based information couched within a narrative reflecting parents' experiences with the condition. This KT tool was adapted (linguistics and imagery) for French- and Tagalog-speaking parents and caregivers through feedback from key stakeholders. The videos were presented to the respective language speakers for usability testing and discussion. Participants were asked to view the KT tool, complete a usability survey, and participate in semistructured interviews. Audio recordings from the interviews were transcribed verbatim, translated into English, and analyzed for relevant themes by using thematic analysis.

Results: French- (n=13) and Tagalog-speaking (n=13) parents completed the usability survey and were interviewed. Although analyzed separately, both data sets produced similar findings, with key themes relating to understanding, relatability, and accessibility. Both the French and Tagalog groups reported that the video and other KT tools were useful in their adapted forms. Participants in both groups cautioned against using verbatim vocabulary and suggested that cultural competency and understanding of health languages were essential for high-quality translations. Parents also discussed their preference for videos with diverse visual representations of families, home environments, and health care workers, as such videos represent their communities more broadly.

Conclusions: French and Filipino parents appreciated having KT tools in their first language; however, they were also supportive of the use of English KT products. Their suggestions for improving the relatability and communication of health messages are important considerations for the development and adaptation of future KT products. Understanding the needs of the intended end users is a crucial first step in producing relevant tools for health evidence dissemination.

XSL•FO
RenderX

Introduction

Background

In the context of pediatric health care, connecting parents and caregivers to research evidence has the power to improve health decision-making and access to support services [1-3]. Knowledge translation (KT) facilitates the actioning of evidence to improve such outcomes. Furthermore, by integrating patient experiences, KT tools (eg, videos, e-books, and infographics) may be more relevant and impactful [4].

Parents and caregivers have become increasingly reliant on web-based sources for health information [5,6]. Subsequently, we developed >25 digital KT tools to improve dissemination and reach of health information for parents seeking help for acute childhood conditions (eg, bronchiolitis, croup, acute otitis media, and fever) [7-11]. The theoretical foundation underpinning our work was the Knowledge to Action framework [12], and the process development for our tools is situated within the tailoring *knowledge* portion of the knowledge to action cycle. Our research teams worked with clinicians and families to co-develop and disseminate the KT tools for Albertans and Canadians in general [13]. Throughout the process, our team engaged with and integrated end user (parent) feedback to ensure that our KT tools were devised and tested with the health consumer in mind [13]. However, our end users have often represented majority cultures in our community, and our KT tools were piloted with and disseminated primarily to English-speaking residents.

Although integrating feedback from end users or stakeholders has become a standard practice for many involved in KT tool development, those providing feedback commonly represent majority voices in the health care community [14,15]. Few studies have explored the usability of KT tools with minority cultural groups [16], and even fewer have tested culturally adapted tools with their target populations [17].

Objective

Our goal was to broaden the reach of our work to different linguistic and cultural contexts to enhance knowledge and awareness among diverse user groups. The objective of this study is to explore how to adapt these tools to 2 non–English-speaking groups. Alberta is a diverse province and continues to grow in represented cultures. Other than English, French and Tagalog are the most common languages spoken in Alberta homes [18]. The relatively high proportion of those who speak French and Tagalog prompted our decision to include these communities in our efforts to understand cultural adaptations for consideration in KT tool development.

To address the different cultural communities in Alberta, we adapted a pre-existing KT tool for French- and Tagalog-speaking parents. By engaging community members, we aim to explore the key cultural aspects of French and Filipino parent experiences as well as to understand considerations for cultural adaptation processes in general.

Methods

The KT Tool

A pre-existing English digital whiteboard animation video depicting the signs and symptoms of croup was chosen for adaptation [18]. This topic was chosen for adaptation as croup is one of the most common causes of upper airway obstruction in children and accounts for significant rates of emergency department visits in Canada [19]. This video was originally produced through consultation with predominantly White, English-speaking parents [20]. Recent usability testing showed that parents rated the tool highly and gave favorable scores to all usability questions. Parents also reported that the video was informative, easy to understand, short and to the point, and visually appealing [21]. Audio and visuals of the original video were representative of mainstream Canadian health care providers' and parents' experiences.

Adaptation Process

Overview

Although the adaptation process for the 2 different cultural groups varied slightly, in the absence of a cultural adaptation theory, model, or framework for KT tools (or KT interventions), we used elements of the Ecological Validity Framework by Bernal et al [22]. Adaptations were considered to be related to language (translation and differences in regional or subcultural groups), persons (patient–health care provider relationship, family roles), metaphors (symbols and sayings), and content (values, customs, and traditions).

French Adaptation Process

The adaptation process for French-speaking parents was purely linguistic. French-speaking members of the research team translated the original English video script into French by using a forward-back translation process [23]. French-speaking researchers and clinicians were consulted during the script translation process to ensure appropriate communication of the medical terminology. A video production company then incorporated the linguistic changes by creating a new French narration and editing visual text to French terminology (Multimedia Appendix 1). Once edits from the digital media company were approved by key stakeholders, the tool was ready for evaluation by the end user group (ie, French-speaking parents).

Tagalog Adaptation Process

The adaptation process for Tagalog-speaking parents involved linguistic and visual adaptation. The linguistic adaptation component of this process was similar to the French adaptation, with narration and video text translation by Tagalog-speaking members of the research team, integration by the digital media company, and review by key stakeholders. We also adapted character appearances to more similarly represent Filipino community members. Through continual consultation with Filipino parent stakeholders and graphic designers from the

digital media company, character visuals were adapted to represent the Filipino community more accurately (Multimedia Appendix 1). Linguistic and visual edits from the digital media company were approved by key stakeholders before evaluation.

Participants and Ethics Approval

Participants were purposively sampled based on self-reported preference for speaking French or Tagalog at home, having English as a second language, and having a parent or guardian role of a child aged <18 years. Ethics approval was granted by the University of Alberta Health Research Ethics Board (Pro00087285 and Pro00085766). All the study documents were translated into French and Tagalog (recruitment materials, study information letters, consent forms, usability surveys, and interview guides). All participants provided informed consent before data collection. Recruitment materials were posted on social media platforms and bulletin boards throughout the respective communities.

Usability Survey

In both studies, parents were asked to complete a usability survey in French or Tagalog after viewing the adapted KT tool. The survey (Multimedia Appendix 2) collected demographic information and evaluated the video's quality of information, format, appropriateness of visuals, and communication of health information. The survey content was informed by a knowledge synthesis of over 180 usability evaluations [24]. Participants were instructed to state their agreement with 9 statements (eg, "it is simple to use" and "its length is appropriate") on a 5-point Likert scale ranging from *strongly disagree* to *strongly agree*. Following survey completion, parents were invited to participate in one-on-one semistructured interviews to elaborate on their survey responses and speak about the usability of the translated tool within their communities.

Interviews

French- and Tagalog-speaking research team members conducted interviews with French and Filipino parents who completed the usability survey. These interviews occurred in French or Tagalog with research team members who were trained in qualitative data collection, following an interview guide and asking relevant probing questions. The interview questions were chosen to explore the participants' perspectives on general cultural considerations, as well as specific feedback for the adapted croup video (Multimedia Appendix 3). The interviews were recorded, transcribed verbatim by a native French- or Tagalog-speaking translator, and then translated into English for analysis. Conducting the interviews in French and Tagalog allowed participants to communicate their perspectives in their preferred language [25]. The decision to translate interview transcripts was appropriate for communicating findings within the predominantly English research team and readership audience. This choice was methodologically consistent for thematic analysis, unlike more deeply phenomenological methodologies in which the participant language could articulate nuances in experience [26]. Nevertheless, any cross-language qualitative study can introduce concerns regarding data trustworthiness [27]. To mitigate potential translation errors, the bilingual research team members validated the translated transcripts and communicated with the broader research team. The English field notes taken by the interviewers also aided in the cross-language study process, where participants' nonverbal responses were noted.

Analysis

Descriptive statistics were used to describe the study sample and the usability survey results. Interview data collection and analysis occurred concurrently until no new responses transpired. The translated transcripts were analyzed using inductive thematic analysis [28]. Data management and analysis were facilitated using NVivo 12 software (version 12; QSR International). The analysis process was iterative, where each transcript was read in its entirety, verbatim codes were assigned to topics in the transcript, and codes were grouped into preliminary categories. The preliminary categories from all the coded transcripts were compared and organized into themes. Common themes were reviewed and verified by all authors. Interviewers wrote field notes during the interviews to promote confirmability, reflexive practice, and rigorous research processes [29]. The trustworthiness of the data was guided by three criteria: fairness, ontological and educative authenticity, and catalytic and tactical authenticities [29]. Fairness was addressed through detailed field notes, interview recordings, and transcripts. Ontological and educative authenticity were addressed through an inductive interview process in which participant perspectives were considered true. Catalytic and tactical authenticities were addressed through continual consultation with key stakeholders, positioning the participants as experts of their own experiences. Analytic rigor was promoted through communication within the research team and the verification of themes. Bilingual research team members validated translated interview transcripts to mitigate interpretation errors that can occur in cross-language qualitative research [27].

Results

Overview

A total of 13 French- and 13 Tagalog-speaking parents residing in Alberta completed the study. The participant demographics are shown in Table 1. Briefly, most French-speaking participants were mothers (12/13, 92%) and were born in Canada (8/13, 62%), whereas Tagalog-speaking participants were mothers (9/13, 69%) and fathers (4/13, 31%) born outside Canada (12/13, 92%).

Table 1. Demographic characteristics of French and Filipino parents who participated in usability testing and interviews (N=26).

Variable	Participants, n (%)	
	French (n=13)	Filipino (n=13)
Parenting role		
Father	1 (8)	4 (31)
Mother	12 (92)	9 (69)
Age (years)		
20-30	2 (15)	1 (8)
31-40	6 (46)	6 (46)
41-50	4 (31)	5 (38)
≥51	1 (8)	1 (8)
Marital status		
Married	10 (77)	13 (100)
Single	3 (23)	0 (0)
Household income (US $)		
<25,000	1 (8)	0 (0)
25,000-49,999	1 (8)	1 (8)
50,000-74,999	1 (8)	2 (15)
75,000-99,999	2 (15)	2 (15)
100,000-149,999	5 (38)	4 (31)
≥150,000	3 (23)	0 (0)
Would rather not say	0 (0)	4 (31)
Highest level of education		
Some high school	1 (8)	0 (0)
High school diploma	0 (0)	0 (0)
Some postsecondary education	0 (0)	1 (8)
Postsecondary certificate or diploma	2 (15)	2 (15)
Postsecondary degree	4 (31)	10 (77)
Graduate degree	6 (46)	0 (0)
Number of children		
1	6 (46)	2 (15)
2	4 (31)	3 (23)
3	1 (8)	8 (62)
4	2 (15)	0 (0)
Born in Canada		
Yes	8 (62)	1 (8)
No	5 (38)	12 (92)

Usability Survey Findings

Overall, parents in both groups found the video helpful, simple to use, and informative. Most French (12/13, 92%) and Filipino (10/13, 77%) parents strongly agreed that the adapted form of the tool was useful. Most parents (10/13, 77% for both French and Filipino parents) also strongly agreed that the tool would help them in making decisions about their child's health (eg, when to use health services and management of the condition). Table 2 displays the detailed responses to the usability survey of the parents in each group.

Table 2. Frequency of responses from French and Filipino parents on usability survey items (N=26).

Items	Strongly agree, n (%)		Agree, n (%)		Not sure, n (%)		Disagree,[a] n (%)	
	French (n=13)	Filipino (n=13)	French (n=13)	Filipino (n=13)	French (n=13)	Filipino (n=13)	French (n=13)	Filipino (n=13)
It is useful	12 (92)	10 (77)	1 (7)	3 (23)	0 (0)	0 (0)	0 (0)	0 (0)
It provides information that is relevant to me	11 (85)	10 (77)	2 (15)	3 (23)	0 (0)	0 (0)	0 (0)	0 (0)
It is simple to use	9 (69)	10 (77)	1 (7)	3 (23)	3 (23)	0 (0)	0 (0)	0 (0)
I can use it without additional help	10 (77)	6 (46)	1 (7)	6 (46)	1 (7)	1 (7)	1 (7)	0 (0)
Its length is appropriate	10 (77)	10 (77)	2 (15)	2 (15)	1 (7)	1 (7)	0 (0)	0 (0)
It is esthetically pleasing	8 (62)	7 (54)	4 (30)	5 (38)	1 (7)	0 (0)	0 (0)	0 (0)
It helps me to make decisions about my child's health	10 (77)	10 (77)	3 (23)	3 (23)	0 (0)	0 (0)	0 (0)	0 (0)
I would use it in the future	5 (38)	10 (77)	5 (38)	3 (23)	2 (15)	0 (0)	1 (7)	0 (0)
I would recommend it to a friend	10 (77)	11 (85)	2 (15)	2 (15)	1 (7)	0 (0)	0 (0)	0 (0)

[a]Options were listed on a 5-point Likert scale ranging from *strongly agree* to *strongly disagree*. The *strongly disagree* option was not chosen by any participant in any category.

Qualitative Interview Findings

Overview

All interviews were one-on-one, with the exception of a Filipino parent dyad who were interviewed together. The analysis of the translated interview transcripts identified the following three major themes: enhancing understanding, relatability, and accessibility. Table 3 displays detailed examples of themes, subthemes, and related participant quotes. It should be noted that the analysis occurred for each community group separately; between-group differences are explained in more detail throughout this section.

Table 3. Themes, subthemes, and quotes developed from interviews with French and Filipino parents.

Themes and sub-themes	Quotes
Enhancing understanding	
Knowing what to do	• "It is rather easy to confuse it with a cold, but you notice a dry and hoarse cough, a noisy breathing." [Interview_004_French] • "I'm even excited to show it and share it to my friends." [Interview_001_Tagalog]
Video format	• "It's excellent, yes very well done, simple, clear, the visual helps a lot, I just like the white and the black, you know it's very well done." [Interview_012_French] • "It seems that the drawing is helpful to get a better understanding of the video." [Interview_013_Tagalog]
Relatability	
Inclusion in Canadian health care	• "I'm more bilingual now than I was before, but when I arrived in 2003, my command of English wasn't as good, and I would have certainly appreciated having this kind of video." [Interview_013_French] • "I feel valued. Just having this, I feel valued already." [Interview_003_Tagalog]
Other child health concerns	• "[my daughter] was diagnosed with pneumonia the last time we went she was diagnosed but they didn't diagnose her with Asthma, they say treating her for Asthma." [Interview_007_French] • "Most of the children have ear infection." [Interview_009_Tagalog]
Video images	• "If you'd like to add something cultural, I'll say you can do it and I'll add something, a little bit of colors or something that symbols but other than that I think it's fine." [Interview_010_French] • "It seems that the drawing is helpful to get a better understanding of the video." [Interview_013_Tagalog]
Accessibility of information	
N/A[a]	• "I don't know where but it will be accessible, but if it's accessible through social networks, through Youtube there are so many videos, who does anything." [Interview_006_French] • "Usually, the method of delivery is when you go to the hospital or your physician doctor, you get a pamphlet, like that. So this is something that [is] a little bit different, and easily accessible anywhere and you don't have to really dig for it." [Interview_012_Tagalog]

[a]N/A: not applicable.

Theme 1: Enhancing Understanding

Subtheme 1: Knowing What to Do

The main goal of the KT tool presented in this study was to help parents understand how to respond when their child has symptoms consistent with croup. The applicability of the tool to French and Tagalog parents could be assessed through the parents' gained understanding of the health condition. Participants from both communities commented that they felt more equipped to respond to future occasions on which their children could have croup. It should be mentioned that although none of the parents had experience with their child having croup, several parents did have experience with their child having laryngitis or asthma. One French-speaking parent described noticing their child had "the barking coughs, with the difficulty to breathe"[Interview_002_French], so they felt prepared to handle a similar symptom in the future. Parents also voiced their appreciation for the health information that they could pass along to friends and family. A Tagalog-speaking participant said the following:

> *I'm even excited to show it and share it to my friends.*
> [Interview_001_Tagalog]

In both the French and Tagalog translated videos, participants were confused by some of the terminology used. The names of drugs and symptom descriptions were the most commonly confused terms in the video, which were terms without verbatim translations. Parents suggested using sounds and visuals to describe symptoms more clearly. Parents also suggested displaying English words for drug names and health conditions to assist the viewer in future conversations with English-speaking health care providers.

Subtheme 2: Video Format

Although parents were presented with a whiteboard animation video for discussion, they were also asked to comment on their preferences regarding the format of KT tools in general. Throughout the interviews, French- and Tagalog-speaking parents described their preferred format as succinct informational videos in simple language. A Tagalog-speaking participant described their positive opinion of the whiteboard animation style in saying the following:

> *It seems that the drawing is helpful to get a better understanding of the video.* [Interview_013_Tagalog]

Theme 2: Relatability

Subtheme 1: Inclusion in Canadian Health Care

Overall, the interviewed French- and Tagalog-speaking parents were pleased with the opportunity to comment on the cultural relevance of the video. Parents liked seeing their languages represented despite most having a high English language

proficiency. A Tagalog-speaking parent voiced this appreciation by saying the following:

> I feel valued. Just having this, I feel valued already. [Interview_003_Tagalog]

Subtheme 2: Other Child Health Concerns

Along with commenting on the present video, French- and Tagalog-speaking parents were asked which child health topics they would like to see represented in future KT efforts. French-speaking parents only mentioned the inclusion of information on asthma and other respiratory conditions. Approximately 15% (2/13) of Tagalog-speaking parents wanted more information tools for eczema, whereas others suggested that juvenile diabetes and treatment of burns would improve their knowledge base.

Subtheme 3: Video Images

Several Tagalog-speaking parents believed that the video images were already representative of their experiences, as their culture was assimilated into a Western Canadian lifestyle. A participant said the following:

> we are more westernized than any other Asians that's why to translate the video in Tagalog to give out information with regard to health care and with the Canadian setting it's really awesome. [Interview_005_Tagalog]

Others viewing the Tagalog adaptation suggested the inclusion of images of religious items in the household to represent the importance of Catholicism for many Filipino community members. Both French- and Tagalog-speaking parents commented on showing more diverse family representations. French-speaking parents suggested that showing a father in addition to a mother would be helpful, whereas several Tagalog-speaking parents thought the inclusion of grandparents would more accurately depict their experiences. A French-speaking parent mentioned the practical considerations for diverse representation in saying the following:

> ...French Canadian, another one from West Africa and then another one from North Africa, and then in fact, heterosexuals or homosexuals, there's lots of possible combinations, I don't think in 3 minutes you can put it all together. [Interview_013_French]

Of note, several parents from both groups situated their experiences as not wholly representative of everyone in their community.

Theme 3: Accessibility of Information

Parents were interested in discussing how to improve the accessibility of information to others in their communities and in Canada in general. Their discussions revolved around the format and dissemination avenues. As previously mentioned, parents strongly preferred using videos as a medium for KT tools. Parents described their previous experiences with finding information as challenging and preferred having easy to access web-based resources. A Tagalog-speaking parent described the importance of web-based tools by saying the following:

> Usually, the method of delivery is when you go to the hospital or your physician doctor, you get a pamphlet, like that. So this is something that [is] a little bit different, and easily accessible anywhere and you don't have to really dig for it. [Interview_012_Tagalog]

Parents also offered suggestions for where to find this information in the future. Many recommended social media sites where parents are already active. A French-speaking parent suggested that the tools should be "accessible through social networks, through Youtube" [Interview_006_French]. Regardless of their recommendations for further adaptation, their preferences for web-based tools and videos similar to the shown croup video were noticeable.

Discussion

Principal Findings

One of the overall goals of our research program is to understand how best to develop and adapt KT tools that are relevant for culturally and linguistically diverse populations. As a first step to support the understanding of best practices when adapting KT products, we adapted a whiteboard animation video for French and Tagalog speakers and sought the perspectives of parents from the respective groups.

Increasingly diverse populations in Alberta and Canada [18] offer new lived experiences and perspectives through which health information can be understood. Efforts to culturally adapt health promotion campaigns [30,31], health care screening inventories [32], and health interventions [33,34] have previously been reported. However, to date, no substantive guiding protocols exist for culturally adapting KT products. As researchers become more interested and able to adapt their KT products to diverse audiences, general considerations will provide guidance in their adaptation endeavors.

By involving stakeholders and engaging end users, we were able to learn some key considerations when adapting tools, which could prove helpful in future KT development or adaptation work. Participants emphasized the importance of enhancing understanding through relatable and accessible KT tool adaptations. French and Filipino community members suggested that future tools should have translations available; primarily use video format; use visuals representative of the community they serve; and be disseminated through commonly used platforms to improve the understanding, relatability, and accessibility of KT tools. On the basis of the recommendations from parents in this study, researchers interested in reaching diverse communities with their KT efforts should consider how best to enhance understanding, relatability, and accessibility within the community of interest. Although these studies involved French and Filipino parents in Alberta, many of the findings could be extended to best practices when engaging with other communities in the KT process. Addressing these areas will undoubtedly look quite different for each specific community of interest; therefore, further patient-centered research should include diverse perspectives.

It is important to acknowledge that culture is not a fixed set of characteristics limited to race and ethnicity; rather, it is a constantly evolving and dynamic concept that encompasses collective views, beliefs, norms, expectations, traditions, customs, and interactions that distinguish population groups [2,3,35]. Therefore, navigating cultural adaptation work can be an involved and complicated process. As mentioned by both Filipino and French parents in this study, understanding the Canadian health care system and Canadian culture more broadly improves with time. Francophone parents were primarily born in Canada; however, Filipino parents who were born elsewhere had an added barrier to accessing Canadian health care services. However, both the French and Filipino parents described that their communities were quite integrated into mainstream Canadian culture; thus, they felt represented by images developed for most cultures. When adapting or developing knowledge products, it is important to assess how the target community has integrated into mainstream Canadian culture. It is likely that newcomers to Canada would experience unique and difficult cultural barriers to accessing health services. Further research into newcomer perspectives in seeking care and how cultural assimilation may affect health information–sharing efforts would add value to the field of KT.

Previous efforts to culturally adapt interventions and health promotion materials suggest that engaging with stakeholders in specific communities will produce more impactful tools for knowledge users [31,32,36]. Although very little data are available on how this is implemented in evidence-based KT tools, public health messaging campaigns have explored the cultural nuances that influence knowledge uptake and interpretation [16,30]. There are a few examples of cultural and linguistic adaptations of health promotion materials for specific cultural communities available [31,36]. Telenta et al [31] found that community member engagement allowed for the successful adaptation of a public health campaign for African migrants in Australia. In addition, Bronheim et al [36] further stressed the importance of involving key community partners to assure cultural and linguistic appropriateness in health campaign development and adaptation. Through these examples, researchers emphasize the importance of engaging with specific cultural groups during the creation of health promotion materials [17,31,36] and acknowledging information preferences to ensure effective KT efforts. Despite the recommendation for co-designed KT products, there is little clarity on the protocols for cultural adaptation co-design research.

When adapting or developing KT products for a particular community, there will undoubtedly be unique considerations specific to that cultural group; however, common threads for researcher practice may exist to guide future efforts. Exploring the effective processes that researchers have used during their cultural adaptation work may provide guidance to streamline the adaptation process for others. In addition, crucial for KT tool development is researchers' willingness to understand the needs of the target community to produce effective tools for health evidence dissemination.

Adapting and translating KT products is a time-consuming and costly endeavor for researchers and may not be achievable for all groups. Certain components of the adaptation process may be prioritized by evaluating the feasibility and impact for the target community. By sharing cultural and linguistic translation processes, we hope to provide guidance for future research efforts in this area.

Limitations and Future Directions

We acknowledge that this study involved participants from only 2 cultural communities in Alberta, Canada; thus, the findings may not be generalizable to the experiences of those from other cultural groups or in other settings and countries. However, as mentioned by the participants in this study, French and Filipino people's experiences differ, particularly in terms of access to support. This suggests that cross-cultural adaptation needs may transcend language and culture and be related to how different community groups assimilate, including their awareness of health care support and how to access them. Therefore, understanding the needs of a community and the intricacies of culture may not be fully available to those outside the community.

The process of adapting KT tools for culturally and linguistically diverse communities can be time consuming and involved. Therefore, studies exploring what elements are cross-cutting versus unique to different communities, as well as what elements are related to culture specifically versus familiarity with the local language and system, are warranted.

In addition, future research understanding the needs of other cultural communities using the engagement of families and community leaders and in collaboration with health care providers would add to this nuanced field of KT.

Conclusions

French- and Tagalog-speaking parents were supportive of the use of English KT products with westernized images but suggested considering cultural components when adapting or developing KT tools. Producing tools with high-quality language translations, video formats, and appropriate and diverse visuals and disseminating on the web would improve understanding, relatability, and accessibility for French- and Tagalog-speaking parents.

As researchers increase their KT efforts, there is a need for more patient-centered cultural adaptation research to inform the development of relevant tools for diverse communities. Important factors such as the community's integration into mainstream culture, literacy, and newcomer status may influence the adaptation process.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1
Adapted knowledge translation tools.
[DOCX File , 554 KB - formative_v6i3e33156_app1.docx]

Multimedia Appendix 2
Usability survey.
[DOCX File , 19 KB - formative_v6i3e33156_app2.docx]

Multimedia Appendix 3
Semistructured interview guide.
[DOCX File , 16 KB - formative_v6i3e33156_app3.docx]

References

1. Albrecht L, Scott SD, Hartling L. Knowledge translation tools for parents on child health topics: a scoping review. BMC Health Serv Res 2017;17(1):1-12 [FREE Full text] [doi: 10.1186/s12913-017-2632-2] [Medline: 28962637]
2. Archibald M, Scott S, Hartling L. Mapping the waters: a scoping review of the use of visual arts in pediatric populations with health conditions. Arts Health 2014;6(1):5-23. [doi: 10.1080/17533015.2012.759980]
3. Graham ID, Tetroe J. How to translate health research knowledge into effective healthcare action. Healthc Q 2007;10(3):20-22 [FREE Full text] [doi: 10.12927/hcq..18919] [Medline: 17632905]
4. Meherali S, Hartling L, Scott SD. Cultural adaptation of digital knowledge translation tools for acute otitis media in low- to middle-income countries: mixed methods usability study. JMIR Form Res 2021;5(1):e13908 [FREE Full text] [doi: 10.2196/13908] [Medline: 33470942]
5. Bouche G, Migeot V. Parental use of the internet to seek health information and primary care utilisation for their child: a cross-sectional study. BMC Public Health 2008;8:300 [FREE Full text] [doi: 10.1186/1471-2458-8-300] [Medline: 18755029]
6. Khoo K, Bolt P, Babl FE, Jury S, Goldman RD. Health information seeking by parents in the internet age. J Paediatr Child Health 2008;44(7-8):419-423. [doi: 10.1111/j.1440-1754.2008.01322.x] [Medline: 18564080]
7. Translating Evidence in Child Health to Enhance Outcomes. Parent Resources – Tools. EchoKT. URL: https://www.echokt.ca/tools/ [accessed 2021-04-15]
8. Le A, Hartling L, Scott SD. The development and usability testing of digital knowledge translation tools for parents of children with bronchiolitis. medRxiv. Preprint posted online June 08, 2021. [doi: 10.1101/2021.06.21.21259266]
9. Le A, Hartling L, Scott SD. The development and usability testing of digital knowledge translation tools for parents of children with acute otitis media. medRxiv. Preprint posted online June 08, 2021. [doi: 10.1101/2021.06.29.21259431]
10. Scott SD, Cunningham C, Le A, Hartling L. Development and usability testing of two arts-based knowledge translation tools for parents about pediatric fever. medRxiv. Preprint posted online June 08, 2021. [doi: 10.1101/2021.06.08.21258574]
11. Scott SD, Le A, Hartling L. Developing and testing an arts-based, digital knowledge translation tool for parents about childhood croup. medRxiv. Preprint posted online June 03, 2021. [doi: 10.1101/2021.06.03.21257424]
12. Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? J Contin Educ Health Prof 2006;26(1):13-24. [doi: 10.1002/chp.47] [Medline: 16557505]
13. Hartling L, Elliott SA, Buckreus K, Leung J, Scott SD. Development and evaluation of a parent advisory group to inform a research program for knowledge translation in child health. Res Involv Engagem 2021;7(1):38 [FREE Full text] [doi: 10.1186/s40900-021-00280-3] [Medline: 34099069]
14. Kothari A, Wathen CN. A critical second look at integrated knowledge translation. Health Policy 2013;109(2):187-191. [doi: 10.1016/j.healthpol.2012.11.004] [Medline: 23228520]
15. Kothari A, Wathen CN. Integrated knowledge translation: digging deeper, moving forward. J Epidemiol Community Health 2017;71(6):619-623. [doi: 10.1136/jech-2016-208490] [Medline: 28298415]
16. O'Brien J, Fossey E, Palmer VJ. A scoping review of the use of co-design methods with culturally and linguistically diverse communities to improve or adapt mental health services. Health Soc Care Community 2021;29(1):1-17. [doi: 10.1111/hsc.13105] [Medline: 32686881]

17. Arora A, Maharaj R, Naidu S, Chimoriya R, Bhole S, Nash S, et al. Views of Indian migrants on adaptation of child oral health leaflets: a qualitative study. Children (Basel) 2021;8(1):28 [FREE Full text] [doi: 10.3390/children8010028] [Medline: 33430177]

18. Turcotte M. Insights on Canadian society: results from the 2016 census: education and labour market successes and challenges for children of immigrant parents. Statistics Canada. 2019. URL: https://www150.statcan.gc.ca/n1/pub/75-006-x/2019001/article/00016-eng.htm [accessed 2021-04-15]

19. Rosychuk RJ, Klassen TP, Metes D, Voaklander DC, Senthilselvan A, Rowe BH. Croup presentations to emergency departments in Alberta, Canada: a large population-based study. Pediatr Pulmonol 2010;45(1):83-91. [doi: 10.1002/ppul.21162] [Medline: 19953656]

20. Hartling L, Scott S, Pandya R, Johnson D, Bishop T, Klassen TP. Storytelling as a communication tool for health consumers: development of an intervention for parents of children with croup. Stories to communicate health information. BMC Pediatr 2010;10:64 [FREE Full text] [doi: 10.1186/1471-2431-10-64] [Medline: 20813044]

21. Scott SD, Le A, Hartling L. Developing and testing an arts-based, digital knowledge translation tool for parents about childhood croup. medRxiv. Preprint posted online on June 18, 2021. [FREE Full text] [doi: 10.1101/2021.06.03.21257424]

22. Bernal G, Bonilla J, Bellido C. Ecological validity and cultural sensitivity for outcome research: issues for the cultural adaptation and development of psychosocial treatments with Hispanics. J Abnorm Child Psychol 1995;23(1):67-82. [doi: 10.1007/BF01447045] [Medline: 7759675]

23. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, ISPOR Task Force for Translation and Cultural Adaptation. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. Value Health 2005;8(2):94-104 [FREE Full text] [doi: 10.1111/j.1524-4733.2005.04054.x] [Medline: 15804318]

24. Hornbæk K. Current practice in measuring usability: challenges to usability studies and research. Int J Hum Comput Stud 2006;64(2):79-102. [doi: 10.1016/j.ijhcs.2005.06.002]

25. Twinn S. An exploratory study examining the influence of translation on the validity and reliability of qualitative data in nursing research. J Adv Nurs 1997;26(2):418-423. [doi: 10.1046/j.1365-2648.1997.1997026418.x] [Medline: 9292378]

26. Squires A. Methodological challenges in cross-language qualitative research: a research review. Int J Nurs Stud 2009;46(2):277-287 [FREE Full text] [doi: 10.1016/j.ijnurstu.2008.08.006] [Medline: 18789799]

27. Squires A. Language barriers and qualitative nursing research: methodological considerations. Int Nurs Rev 2008;55(3):265-273 [FREE Full text] [doi: 10.1111/j.1466-7657.2008.00652.x] [Medline: 19522941]

28. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

29. Lincoln YS, Guba EG, Pilotta JJ. Establishing trustworthiness. In: Lincoln YS, Guba EG, Pilotta JJ, editors. Naturalistic inquiry. Beverly Hills, CA: Sage Publications; 1985:289-327.

30. Schliemann D, Su TT, Paramasivam D, Somasundaram S, Ibrahim Tamin NS, Dahlui M, et al. The systematic cultural adaptation of a UK public health cancer awareness raising programme for Malaysia: the Be Cancer Alert Campaign. Transl Behav Med 2019;9(6):1087-1099 [FREE Full text] [doi: 10.1093/tbm/ibz134] [Medline: 31583402]

31. Telenta J, Jones SC, Francis KL, Polonsky MJ, Beard J, Renzaho AM. Australian lessons for developing and testing a culturally inclusive health promotion campaign. Health Promot Int 2020;35(2):217-231. [doi: 10.1093/heapro/day118] [Medline: 30809652]

32. Grasaas E, Fegran L, Helseth S, Stinson J, Martinez S, Lalloo C, et al. iCanCope with pain: cultural adaptation and usability testing of a self-management app for adolescents with persistent pain in Norway. JMIR Res Protoc 2019;8(6):e12940 [FREE Full text] [doi: 10.2196/12940] [Medline: 31162132]

33. Matarazzo BB, Hoffberg AS, Clemans TA, Signoracci GM, Simpson GK, Brenner LA. Cross-cultural adaptation of the window to hope: a psychological intervention to reduce hopelessness among U.S. veterans with traumatic brain injury. Brain Inj 2014;28(10):1238-1247. [doi: 10.3109/02699052.2014.916419] [Medline: 24946053]

34. McCool J, Tanielu H, Umali E, Whittaker R. Assessing the cross-cultural adaptation and translation of a text-based mobile smoking cessation program in Samoa (TXTTaofiTapaa): pilot study. JMIR Mhealth Uhealth 2018;6(8):e173 [FREE Full text] [doi: 10.2196/mhealth.9033] [Medline: 30170994]

35. Straus SE, Tetroe J, Graham I. Defining knowledge translation. CMAJ 2009;181(3-4):165-168 [FREE Full text] [doi: 10.1503/cmaj.081229] [Medline: 19620273]

36. Bronheim S, Sockalingam S. A guide to choosing and adapting culturally and linguistically competent health materials. National Center for Cultural Competence, Georgetown University Center for Child and Human Development. 2003. URL: https://nccc.georgetown.edu/documents/Materials_Guide.pdf [accessed 2021-03-30]

Abbreviations

KT: knowledge translation

XSL•FO
RenderX

Original Paper

Adapting an Evidence-Based e-Learning Cognitive Behavioral Therapy Program Into a Mobile App for People Experiencing Gambling-Related Problems: Formative Study

Gayl Humphrey[1], BSc, MSocSc; Joanna Ting Chu[1], PhD; Rebecca Ruwhiu-Collins[2]; Stephanie Erick-Peleti[2], MPH; Nicki Dowling[3], PhD; Stephanie Merkouris[4], PhD; David Newcombe[5], PhD; Simone Rodda[5], PhD; Elsie Ho[5], PhD; Vili Nosa[6], PhD; Varsha Parag[1], BSc; Christopher Bullen[1], MBChB, PhD

[1]National Institute for Health Innovation, University of Auckland, Auckland, New Zealand
[2]Hapai Te Hauora, Auckland, New Zealand
[3]Deakin University, Melbourne, Australia
[4]School of Psychology, Deakin University, Melbourne, Australia
[5]Social and Community Health, University of Auckland, Auckland, New Zealand
[6]Pacific Health, University of Auckland, Auckland, New Zealand

Corresponding Author:
Gayl Humphrey, BSc, MSocSc
National Institute for Health Innovation
University of Auckland
Private Bag 92019
Auckland, 1145
New Zealand
Phone: 64 21 1100901
Email: g.humphrey@auckland.ac.nz

Abstract

Background: Many people who experience harm and problems from gambling do not seek treatment from gambling treatment services because of personal and resource barriers. Mobile health (mHealth) interventions are widely used across diverse health care areas and populations. However, there are few in the gambling harm field, despite their potential as an additional modality for delivering treatment and support.

Objective: This study aims to understand the needs, preferences, and priorities of people experiencing gambling harms and who are potential end users of a cognitive behavioral therapy mHealth intervention to inform design, features, and functions.

Methods: Drawing on a mixed methods approach, we used creators and domain experts to review the GAMBLINGLESS web-based program and convert it into an mHealth prototype. Each module was reviewed against the original evidence base to maintain its intended fidelity and conceptual integrity. Early wireframes, design ideas (look, feel, and function), and content examples were developed to initiate discussions with end users. Using a cocreation process with a young adult, a Māori, and a Pasifika peoples group, all with experiences of problem or harmful gambling, we undertook 6 focus groups: 2 cycles per group. In each focus group, participants identified preferences, features, and functions for inclusion in the final design and content of the mHealth intervention.

Results: Over 3 months, the GAMBLINGLESS web-based intervention was reviewed and remapped from 4 modules to 6. This revised program is based on the principles underpinning the transtheoretical model, in which it is recognized that some end users will be more ready to change than others. Change is a process that unfolds over time, and a nonlinear progression is common. Different intervention pathways were identified to reflect the end users' stage of change. In all, 2 cycles of focus groups were then conducted, with 30 unique participants (13 Māori, 9 Pasifika, and 8 young adults) in the first session and 18 participants (7 Māori, 6 Pasifika, and 5 young adults) in the second session. Prototype examples demonstrably reflected the focus group discussions and ideas, and the features, functions, and designs of the Manaaki app were finalized. Attributes such as personalization, cultural relevance, and positive framing were identified as the key. Congruence of the final app attributes with the conceptual frameworks of the original program was also confirmed.

XSL•FO
RenderX

Conclusions: Those who experience gambling harms may not seek help. Developing and demonstrating the effectiveness of new modalities to provide treatment and support are required. mHealth has the potential to deliver interventions directly to the end user. Weaving the underpinning theory and existing evidence of effective treatment with end-user input into the design and development of mHealth interventions does not guarantee success. However, it provides a foundation for framing the intervention's mechanism, context, and content, and arguably provides a greater chance of demonstrating effectiveness.

Introduction

Background

Gambling problems are viewed across a continuum of risk in New Zealand, ranging from no risk to problem gambling, where high-risk gambling behavior results in considerable health and social problems. Consistent with this framework, the term problem gambling is described as any kind of harm, distress, or adverse impact caused or exacerbated by a person's gambling [1-3]. Gambling-related harm is a significant public health problem in New Zealand [4,5] and internationally [6-8]. In New Zealand, the latest prevalence estimates suggest that 0.1% of the population experiences problem gambling, with a further 1.5% and 3.3% experiencing moderate- and low-risk gambling, respectively [9].

Different treatment service modalities, such as face-to-face counseling, phone support, self-management workbooks, and web-based tools, such as those provided on the Safer Gambling Aotearoa website [10], are readily available, although none are mobile health (mHealth) apps. A recent survey reported that only 3.2% of people with problem gambling had accessed at least one type of support service [9]. Therefore, a significant proportion of the population does not receive optimal therapeutic support [11]. In New Zealand, the Māori (indigenous people of New Zealand) and Pasifika peoples are disproportionately impacted by problem gambling [12,13]. Although some theories of addiction help explain gambling addiction attributes [14], the influence of historical and cultural intergenerational experiences is equally important in understanding the gambling context for Māori [15,16]. Thus, it is vital that interventions accommodate and reflect the cultural appropriateness and presentation of content relevant to their needs and understanding.

The ubiquity of smartphone ownership and the exponential growth of new mobile technology functionality provide an opportunity to add new modalities for delivering evidence-based health interventions [17]. mHealth apps, defined as the use of mobile and wireless technologies for health promotion [18], have extended support for health behavior changes beyond standard treatment contexts. These interventions continue to be in demand, with a reported 3.7 billion mHealth downloads in 2017 [19] and an expected 29% growth rate over 2020-2027, equating to a forecast of 311.9 billion by 2027 [20]. Nonetheless, the retention and use of mobile apps is less than ideal, with reports suggesting that most are uninstalled within 5 days, with mHealth apps deleted slightly later, at 8.8 days [21]. However, app reinstalls are also not unusual, with mHealth apps reporting a reinstall rate of 14% [21].

Some mHealth research reviews explain that the lack of underpinning theory to support the app's intent (ie, the set of determinants that influence cognitive and behavior change assertions) contributes to the lack of app retention and hence the paucity of demonstrable evidence of effectiveness [22-27]. Other mHealth research highlights that the lack of involvement of intended end users or consumers in the design and development is an essential yet missing ingredient in creating effective mHealth interventions [28,29].

Therefore, informing mHealth interventions with research evidence, insights from domain experts, and a theoretical framework for high-fidelity interventions are critical [30], as is the involvement of consumers and intended end users [31-34]. Blending these strategies into concrete elements for digital interventions is not straightforward [35]. Although adopting this approach does not guarantee success [36,37], it can optimize the creation of an effective mHealth program. To that end, there has been a call to publish research that outlines the development processes underpinning mHealth interventions [38,39]. Publishing and sharing such research, which involves both mapping theoretical constructs and cocreation with end users, is a positive step in building a robust and replicable evidence base.

Objectives

In this study, we describe the formative research that was an integral part of developing a mobile app for gambling harm for use in New Zealand: Manaaki. Specifically, we aim to (1) adapt and refine an evidence- and web-based program, GamblingLess [40-42], into a mobile app intervention and (2) engage domain experts and intended end users to help inform the content conversion and the design, features, and functions of the app and maximize the effectiveness of the program. This study is part of a larger project to evaluate the effectiveness of this designed cognitive behavioral therapy (CBT) mHealth intervention—Manaaki—for people with self-reported gambling problems [43].

Methods

Ethics Approval

Ethical approval was obtained from the New Zealand Health and Disability Ethics Committee (19/STH/100). As part of the ethics consenting process, participants gave consent for a summary of the focus group contributions, including quotes where relevant, to be included in publications as part of the participant consenting process.

Development Process

The development process encompassed 2 key phases.

Phase 1 consisted of the research team reviewing the current GamblingLess program and adapting and refining it into a first draft prototype of content, features, and functions suitable for a mobile phone–delivered intervention.

Phase 2 involved a qualitative approach using repeated focus groups and prototype testing to engage end users as collaborative design partners. This approach was used to inform and refine the app's features, style and functions, and content presentation.

Phase 1: Intervention Content Development

Overview

The multidisciplinary research team included clinical gambling research domain experts, including GamblingLess [40] program creators (Associate Prof Dowling, Dr Merkouris, and Dr Rodda, Deakin University, Australia); mHealth and digital health experts; public health physicians; Māori, Pacific, and Asian health addiction research experts; a statistician; and 2 researchers from Hāpai te Hauora (an Iwi-led [groupings of the indigenous population based on kinship] community health organization). Collectively, these team members provided expert advice to unpack the program's core elements to ensure that the content remained evidence based. The Aotearoa New Zealand context also informed practical and clinical considerations during the design and development phases.

Unpacking GamblingLess

GamblingLess [42] is a comprehensive and intensive web-based CBT program designed to help people with gambling problems. It was designed to emulate the intensity and depth of the face-to-face programs. The original program content included 4 modules that were delivered over 8 weeks. The modules included motivational enhancement, behavioral strategies, cognitive strategies, and relapse prevention, and each module included 13-15 discrete activities. Although it was recommended that participants complete all modules and activities in numerical order, they were allowed to complete as many activities as they liked, in any order they chose.

Associate Prof Dowling, with the support of Dr Merkouris and Dr Rodda, initially reviewed the 4 modules of GamblingLess, reflecting on the findings from their original research [41]. This revised program intentionally maintained the original GamblingLess principles underpinned by the transtheoretical model [44]. The key was to acknowledge that some end users (help seekers) would be more ready to change than others. As change is a process that unfolds over time, a nonlinear progression is needed, enabling the end users to select the interventions they need according to their stage in the change process.

Accordingly, the goal of this redeveloped program is to support the end user in selecting and using the most appropriate module or modules based on their needs. It is designed to be nonlinear, and hence presents the end user with a range of options (modules) with which one or more resonates with what they hope to achieve; that is, managing urges or relapse prevention.

Therefore, it was anticipated that not all end users would need or want all modules or all activities within the program.

To create Manaaki, the acceptability data from the original GamblingLess program were also reviewed, with the most helpful treatment content identified and assessed for inclusion in this program [41]. The outcome of this revision was that the original 4 treatment modules were expanded to 6 modules. The research team reviewed the 6 new modules and provided initial suggestions on how each module and associated actions or activities might be adapted to fit a smartphone mobile app interface. Iterative cycles of face-to-face meetings, web-based meetings, and email conversations were used to create and refine the core structure, content, and activities of the 6 new modules.

The researchers adopted the consumer engagement strategies common in marketing [45,46], and with a user experience designer and research partners from Hāpai te Hauora, ideas and examples for the draft elements for each module and the control app were drafted. These ideas were translated into the following mixed medium modalities for use during the cocreation phases with end users:

- A draft wireframe to illustrate flow and give depth and perspective using Adobe XD (Adobe Inc)
- Large A0-sized posters of each module's wireframes, including color palette ideas
- A set of A4-sized paper mobile phone screens to show draft content
- A set of various images for app screens

Phase 2: Cocreation With End Users

Overview

We used a qualitative cocreation methodology [47] to determine the look, feel, function, and features of the Manaaki intervention app. As gambling impacts more marginalized and economically deprived populations, the research team created three groups: a Māori group, a Pasifika group, and a specific young adult group that was limited to participants aged 18-25 years. These groups were selected to maximize the app's acceptability for those most likely to benefit the most. In all, 3 cocreation cycles were planned for each group. This approach encourages increased involvement and ownership of the process by participants, as they can see how their contributions help shape each design iteration of the app [48].

Researchers from Hāpai te Hauora led participant recruitment. Through their networks, contact was made with various community groups and organizations to help identify potential participants for the different groups. The eligibility criteria were as follows: (1) having experienced problems or harms from their gambling or supporting someone who is or has experienced problems with gambling, (2) able to understand and converse in English, (3) able to participate in at least two of the three cocreation cycles, and (4) aged ≥18 years for the Māori and Pacifika groups or 18-25 years for the young adult group.

The focus groups were conducted between July and November 2019. The focus groups for the Māori and Pasifika groups were held in Auckland, whereas for the young adult group, it was held in Wellington. Each workshop lasted between 60 and 90

minutes and was audio recorded with notes taken. The workshops were facilitated by the Hāpai te Hauora researchers and were attended by the design experts.

A mixture of methods was used to elicit feedback and comments from participants. These included open-ended questions, iteratively adding to or removing elements on the Adobe XD wireframes, Post-it notes, writing on A0-sized posters, and ad hoc prompts for a more detailed discussion when new ideas emerged.

After each focus group, various engagement modalities (posters and Adobe XD wireframes) were adapted to reflect emerging ideas and discussions. These were incorporated into the next prototype version, ready for participants to reassess, discuss, and modify the subsequent focus group sessions.

As an appreciation of their time, each participant received an NZD $30 (US $20.30) voucher to attend the first focus group and an NZD $20 (US $ 13.50) voucher for each additional focus group attended.

Data Analysis

The data for the focus groups were analyzed using a general inductive approach [49]. The purpose of using an inductive approach was to (1) condense raw textual data into a brief summary format (using images, words, wireframes, and prototypes), (2) establish clear links between the research objectives and the summary findings derived from the raw data, and (3) develop a framework of the underlying structure of experiences or processes that are evident in the raw data. All focus group data were grouped under the relevant themes, unless it was pertinent to describe a particular group's findings separately. All direct quotes are in *italics* and illustrate specific conversation points.

Results

Phase 1: Converting the Internet-Based GamblingLess Program to a Draft mHealth Program

On the basis of the GamblingLess research findings, the original 4 core modules were altered into 6 new modules, each with 5-7 learning topics. Each learning topic had a set of activities or interactions to support the learning intent of the module's theme (Figure 1).

Each new module was carefully reviewed and remapped to the underpinning conceptual frameworks and intent to ensure fidelity and conceptual integrity (Table 1).

The outcome of this process enabled early wireframes, design ideas (look, feel, and function), and content examples to be drafted, which helped initiate discussions and ideas with end users during phase 2 and cocreation with end users.

Figure 1. Mapping web-based GAMBLINGLESS program content to the Manaaki mobile app content. mHealth: mobile health.
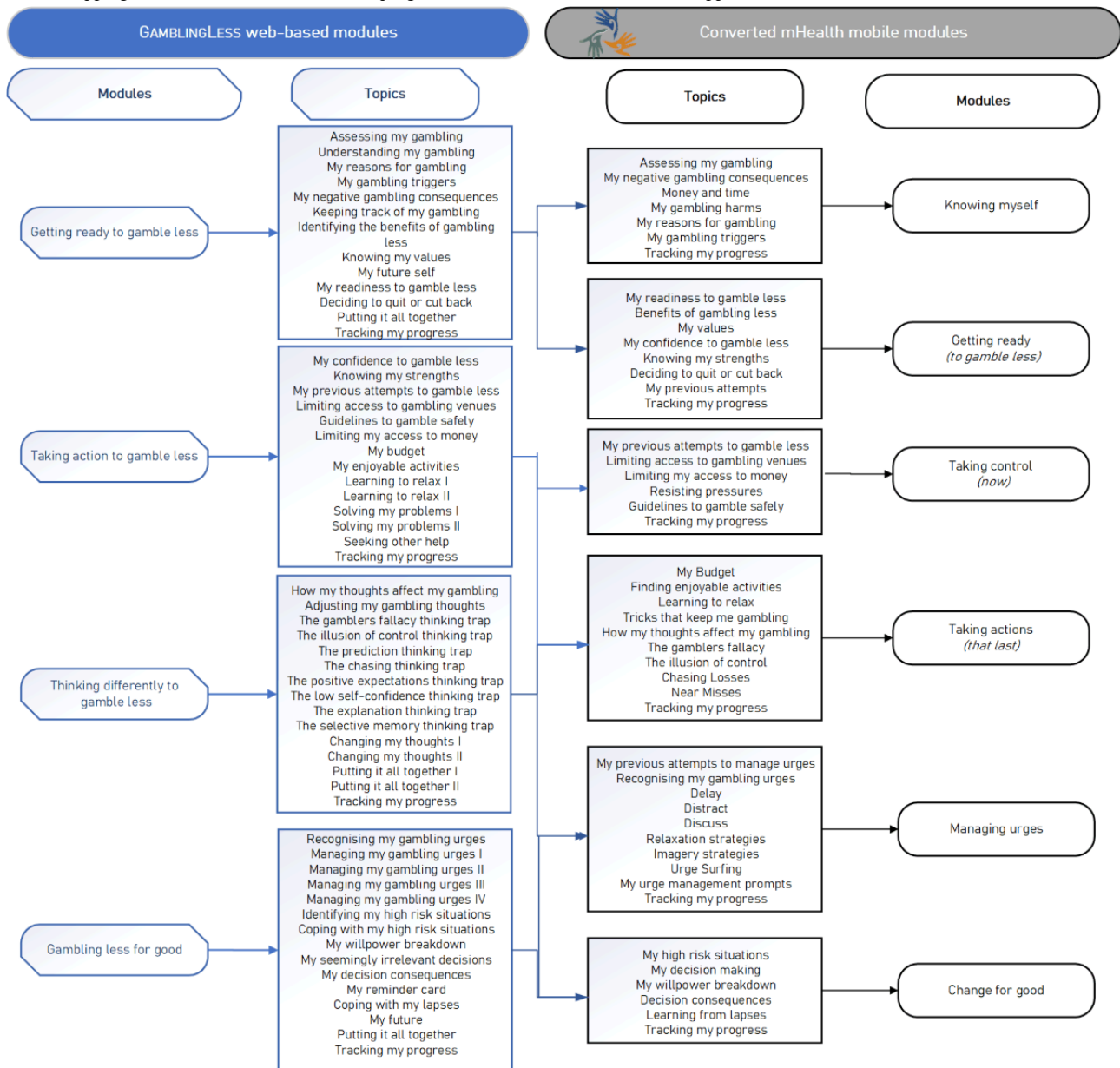
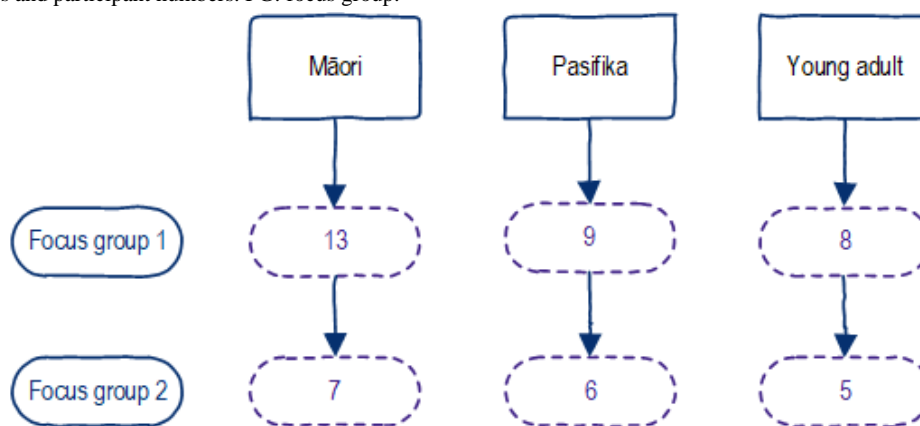**Table 1.** Remapping modules to concepts and the intervention intent.

Module names	Conceptual framework	Underpinning intent
Knowing Myself (and my gambling)	Developing self-awareness and insight [50-52]	Designed to provide personalized feedback, goal setting, and understanding of gambling motivations, triggers, and consequences.
Getting Ready (to make changes)	Targeting thoughts and feelings and activating behaviors [53]	Designed to enhance readiness and confidence to gamble less, helping to shape thoughts and values to help make a change. Incorporates ideas of prediction of outcomes from their behaviors.
Taking Control (right now)	Targeting practical behaviors and identification of situational and contextual triggers	Designed to identify strategies to be used to contain and control the problematic gambling behavior in the short term. Directs the end user to other useful tools, such as venue exclusions, managing money, services and other resources.
Taking Actions (that last)	Activating personal strengths and resources and enhancing belief for successful change [54,55]	Designed to identify strategies and skills to be used to ensure longer-term success in gambling less.
Managing Urges (to cope with real situations)	Reframing thoughts and reflecting on your future self [54,55] and relapse prevention [56] (see also Marlatt and Gordon, 1985)	Designed to help the end user to cope with urges and cravings.
Change for Good (and building a new future)	Relapse prevention [56]	Designed to prevent gambling relapses in the future and to support cognitive capability to understand their lived experience of gambling harms and problems as a mechanism for maintaining their new behavior.

Phase 2: Cocreating and Designing With End Users

Overview

In all, 2 rounds of focus groups were conducted, with 30 unique participants (13 Māori, 9 Pasifika, and 8 young adults) in the first session and 18 participants (7 Māori, 6 Pasifika, and 5 young adults) in the second session (Figure 2). The time between the 2 focus group sessions varied from 4 to 7 weeks among the groups, as arranging an agreeable time for all participants was difficult.

Figure 2. Focus groups and participant numbers. FG: focus group.

Focus Group Cycle 1

Gathering Experiences

Participants were encouraged to think aloud and share difficulties when they or a loved one experienced gambling problems. The concept of *whakama* or shame was the dominant theme across all groups. Some participants remarked that they did not want to acknowledge that they had a problem with their gambling and had a strong desire to keep their gambling problems a secret. Consequently, few reported seeking professional help, but most remarked that they had completed a web-based quiz that assessed their gambling. Even when the quiz results suggested that they may have had a significant gambling problem, taking action was self-reported as trying the strategies that the websites had recommended. Although most remarked that these strategies were helpful, many found that they had returned to their previous harmful gambling behavior. A participant remarked that they felt that without support or reminders, it was easy to forget the learnings. The idea of a mobile phone intervention (*app*) that could provide help wherever and whenever needed was positively established across all 3 groups.

Initially, all groups discussed and conceptualized the app as a one-off tool to be used and then discarded it. However, as the participants continued to share their experiences and discuss what they would find helpful, the concept of a one-off tool was abandoned. The app was increasingly described as something that would be used to *help now and later*. Terms such as
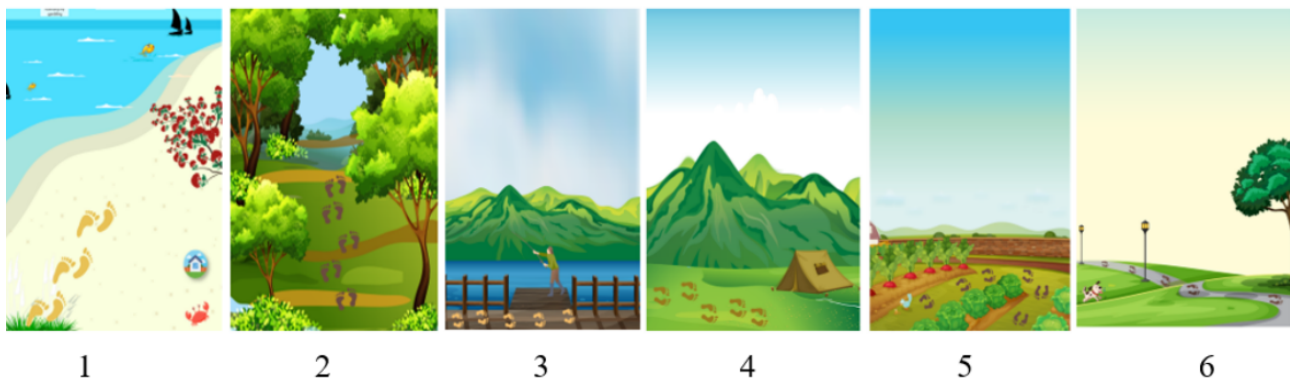
faigamalaga (Pacific group) or *haerenga* (Māori group) were used to convey the idea of recovery and about being on a journey or voyage. The Māori focus group also discussed the concept of *kotahitanga* (togetherness and sharing) as an important concept and that recovery with whānau (family or extended family) is crucial, as it is "through our whānau āwhina" (extended family support) "that we can be and stay well."

The Look and Feel of the Intervention App

In all, 12 draft design images were presented to the participants, and it was clear that the participants expressed excitement in seeing their ideas of a journey already coming to life. As each focus group began selecting images that they liked, conversations within the groups became energetic. Although all 3 groups navigated these discussions differently, their outcomes were similar. Each group concluded that keeping a range of different images of New Zealand was better than keeping a single image or theme.

The 6 images that were the most popular across all 3 focus groups are presented in Figure 3.

In addition to selecting these images, engagement with the images was such that participants thought of ideas to enhance their look and feel, such as to "add more birds; have the bush image a bit more open" (Figure 3; image 2). The footprints were designed to illustrate movement; yet a Pasifika participant wanted to "remove the footprints," whereas others in the group liked them. The footprints were not mentioned by the young adult or Māori group participants. The fisherperson and tent in Figure 3 (images 3 and 4, respectively) were disliked across all 3 groups, commenting that they "did not fit with the other images." A few participants in the Māori focus group remarked that the perspective in these 2 images was different from the others, and "it didn't work." The young adult group remarked that Figure 3 (image 6) "needs people and needs some life in it. It's too sad looking, aye." Similarly, the Pasifika group said Figure 3 (image 6) "looks a bit too lonely. Where is everyone?" The Māori group participants also commented that Figure 3 (image 6) needed "people and activity in that scene so that you felt part of the world, you know, just being normal with your whānau. Yeah, more life."

Figure 3. The 6 most popular initial concept imagery selected.

Shaping the Features, Functions, and Content Desired

The conversations naturally segued from the look and appeal of the images to the important attributes (features and functions) needed to make the app useful and easy to use. The participants in all 3 focus groups began noting down the key features, functions, and some content elements they wanted in the app, using Post-it notes or writing directly on the posters, Textbox 1 summarizes these initial attributes.

Textbox 1. Early attributes to have in the app from the all first focus group sessions.

Features

- Make it personal
- Able to share with others
- Progress and achievement
- Receive feedback
- Measuring where I am
- Self-checks
- Help guide me to where to go next
- Reminders and prompts
- Links to services if needed
- Tracking on how I am going
- Goals
- A place for reflection and celebration
- Quick access to other help if needed

Functionality

- Easy to use
- Navigation around makes sense
- No need to do everything
- Be interactive
- Short things to do
- Not too many clicks to find things
- Needs to be fun and have animation
- Quick, no lagging when loading

Content

- Simple language
- Use images
- Video
- Not too much text
- Positively framed
- Can listen not just read
- Self-help resources
- What other help is available
- Culturally appropriate
- Shows helpful skills
- Activities

Cross-cutting elements

- Secure, private, and confidential
- Easy to download, not complicated
- No own data use
- Not too intruding
- No cost
- Works offline

Creating Ownership of the App Name and Logo

In consultation with our Māori research team members, the name *Manaaki* was suggested as a possible name for the mHealth intervention app. The name Manaaki has a special meaning in Māori, which blends the words mana and āki. Mana means to bring strength or give strength and to cherish. Almost every activity is linked to the maintenance and enhancement of mana. The word āki encompasses the concept of encouragement or urge for support, as well as to hold fast and be strong. To ensure that this term was acceptable and appropriate, the team sought guidance from the governance group at Hāpai te Hauora (our Māori research partner organization). They supported the use of the word Manaaki as the name for the new app.

The focus group participants had positive remarks about the name Manaaki. "I love the name Manaaki; it is perfect," said a Māori group participant. "It gives a positive vibe," was reported by a young adult group participant, and a Pasifika participant remarked, "I like it. Manaaki is all about being encouraging, so perfect for this app."

In all, 10 concept designs (images, colors, and text types) were created, and participants were encouraged to comment on the options provided and select the text, logo, and colors they liked or offered other ideas.

Figure 4 shows the text, images, and colors that were frequently selected, with those circled in red confirmed as the final designs.

Figure 4. Concept designs for name and logo.

Encouraging Recruitment Into the Study

The final discussion element for the first focus group was gathering insights into the study design—a randomized waitlist control study—and its impact on recruiting participants. It was evident that participants in all the groups understood that tools and treatments offered to people experiencing harm from gambling needed to be effective and that research is needed to prove effectiveness. Despite understanding the need to research the app for effectiveness, both the young adult group and the Pasifika group remarked, "that having to wait to get the full app is a put-off." Similar remarks were made in the Māori group:

> If you have to wait, you need to let us know that we can still go look for help. You know, that being in this study thing, doesn't stop that. Gotta be real clear too, about the information saying that, all the stuff in the app will be able to be used, after the wait. [Māori focus group, P3]

Although the idea of waiting was viewed as negative, having some functions was suggested as a way to help retention. The following comment clearly reflects this:

> If I had made the decision to do something about my gambling, and then was told, oh wait for a bit. [Laugh] Well, I would think, that's useless, and delete it and try something else. It may even put me off. But you know, seeing that there is still something for me, well maybe, [Laugh], that's better than nothing. So

> I might stick with it a bit, who knows. [Young adult participant, P2]

Functions and attributes commonly mentioned across the groups had a countdown icon showing how long to when the full Manaaki app would be available. The young adult group also suggested that the countdown icon be a clickable button that could *flip* and show the same outcome using different metrics, such as months, weeks, or days. Common to all groups was that the waitlist app needed to offer support within the app. Support such as links to services or websites was frequently noted, with the Pasifika group also mentioning that links to money management would also be good to have.

Focus Group Cycle 2

The purpose of the second focus group was to present the changes made to the designs and show the features and functions that were added or modified based on the group contributions from the focus group 1.

Look and Feel

The new image designs (Figure 5) and an animated version using Adobe XD were presented to each focus group, and they all received positive comments such as, "These are great. I love the colours, and the animation is sweet," from a young adult participant.

All groups remarked that the various interactive content and features that support users engaging with and using the app

were clear and intuitive. The use of different styles (images and themes) for different modules was received positively (Figure 6).

An animated Tui (a native New Zealand bird [Figure 7]) was designed to appear at the beginning of each module to explain the learning intention and topics for the module. This information was presented in an auto-scrolling speech bubble next to Tui. The participants could also listen to the message using the audio option. This audio feature was highly liked and recommended by all the focus group participants, although a

few comments from the Māori focus group remarked that the "computer voice is a bit annoying." In contrast, some participants in the Pasifika focus group were excited by this option and thought this audio option could enable the text to be spoken in different languages.

Participants were noticeably excited when they saw how the text, images, and colors that were selected during their first focus group session formed the framework for the final branding of Manaaki.

Figure 5. Modified screen imagery based on participant feedback.
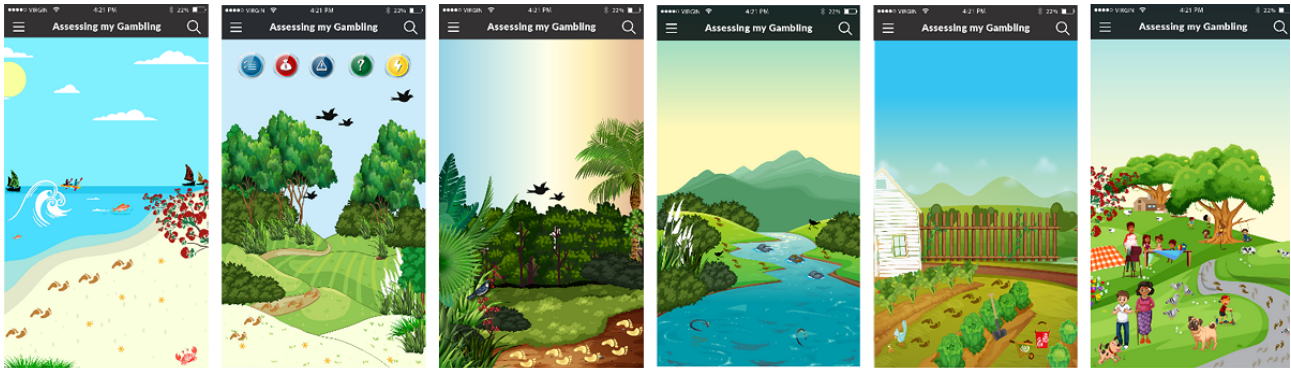
Figure 6. Example of ideas for presenting text and activity interactions.

Example of text presentation by using tabs

Example activity with choices scrolling in from side and fading if selected

Alternate activity selection ideas with selections shown at the bottom
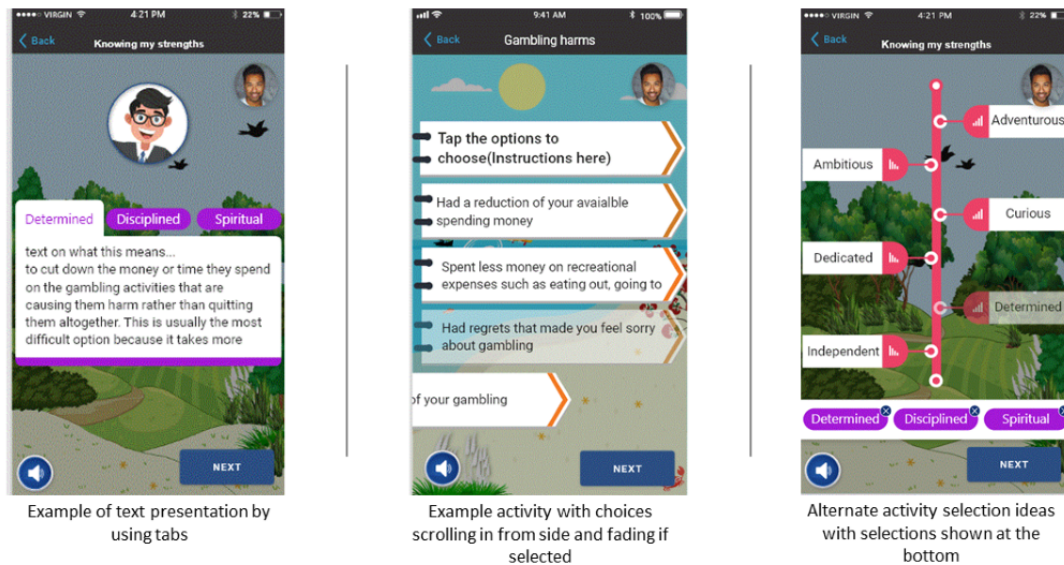
Figure 7. Tui guide.
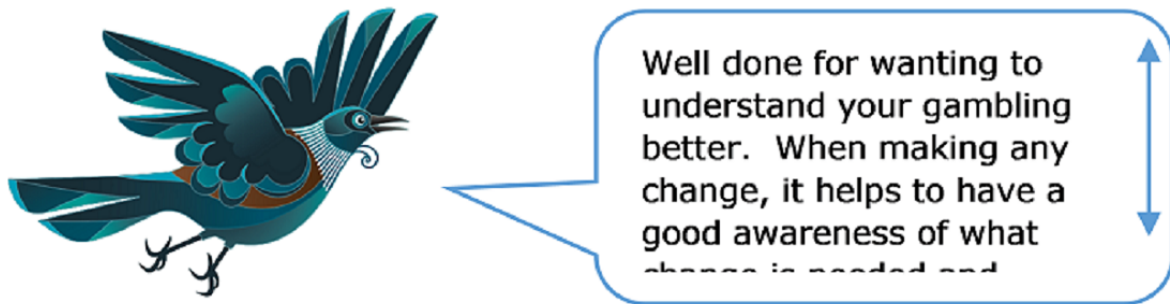
Personalization

In focus group cycle 1, all participants expressed a strong desire for the app to be personalized. The examples that showed how a user's first name would be integrated through the app were highly commended. Similarly, the use of language-specific greetings and phrases throughout the text was seen as acknowledging the population Manaaki intended to support. For many, these "little things [that will] help keep people engaged." Similar comments were made when the function of uploading an image or selecting a personalized avatar to represent a user visually was presented as an idea for how the app could be personalized to the user.

Participants also acknowledged that people could be at different stages of their journey to reduce the behaviors contributing to their gambling-related harm. They all remarked that the designs reflected their ideas from the first focus group sessions, especially the way the app was designed to enable the user to navigate their journey and yet receive prompts and suggestions to help guide them. Other functions, such as the journal, the link to the helpline chat, and the services being visible using the maps, were remarked on as meeting different users' needs and stages, making Manaaki useful to a diverse range of potential people.

Positivity

Positive framing and strength-based content were considered important, and the use of positive and simple language was repeatedly remarked on as being successfully shown across all focus groups:

> *I like how it is about knowing strengths.* [Pasifika focus group, P6]

Although some changes still such as the following needed to be made:

> *Some words are more researchy, like - Fallacy - wouldn't be understood by most. Use a simple way to say this.* [Māori focus group, P1]

> *Maybe instead of looking at what are you spending, it could be, umm, learning to save?* [Young adult focus group, P2]

The range of notification message examples was presented as suggested in the first focus groups as an additional way to provide support and insights to a user as they progressed through activities in the app. These were positively remarked upon, and the young adult groups suggested that a mix of notifications and pop-ups would be more appealing than a single modality. The motivating and encouraging message examples "hit the spot," said a young adult participant. All groups liked the notifications that prompted people to use the app, and many felt that this functionality would help with engagement and retention.

Ability to Share Insights and Progress With Others

The participants suggested that the app could share progress or success with others. This type of networking system was suggested as a positive method for users to encourage each other, share tips and strategies, and provide a sense of *not being alone*. This feature is technically complex, and as such, the solution presented to the groups was that sharing would be limited to those within the study. Feedback from all 3 groups indicated that they would have liked to connect with others beyond the study, as these broader social connections were important.

Tracking My Progress

Being able to track progress appealed to participants in focus group cycle 1. In total, 2 examples were presented as a way to represent progress. These were progress bars under the modules and the emergence of footprints on each module's home screen. The progress bars were identified as the best across all groups, primarily because they were used in other apps and as such, were familiar and easy to understand.

In several instances, participants in the Pasifika group indicated that the app should be able to track the amount of money spent on gambling, money saved from not gambling, and how much time they had or had not spent gambling. Various ideas have been suggested for presenting this in the app. The design that resonated across all groups was one in which the time and money recorded as spent gambling was converted into other activities; for example, the number of dinners with family or friends or movies went to with friends.

Use Images and Multimedia

The app interface needed to be engaging and attractive to all the participants. The young adult group remarked that if the app looked unattractive at the start, they were unlikely to engage. Updated high-fidelity images were shown to the groups, and the following remark summarizes the generally positive comments made across all the groups:

> *Good background now that there is Pacific/Māori punch to it – waka, marae, native birds. Thanks.* [Māori focus group, P11]

Regarding the inclusion of multimedia, feedback was mixed, with participants in the Māori and Pasifika groups liking that short videos from real gamblers were part of the app. The young adults group participants were less enthusiastic, but they liked the cartoon animation videos than those with real people. Having different video styles was seen as a good option.

Overwhelmingly, all participants were concerned that having these videos in the app might affect the ability to download the app, that is, too large or that it would take up too much of their phone's storage, slowing down their phone or stopping other functionality. Hearing that the videos would not be embedded within the app but would be stored externally and linked to YouTube generated positive remarks.

Connect With Treatment Services

The participants in all groups agreed that connecting to a professional or treatment provider through the app would be

advantageous. Participants were shown the services feature and interactive function that presented services by geography (map). List-based search options (Figure 8) were good options as illustrated by the following comments:

> *This feature is perfect, it allows users to look for help without needing to know the name or address of the place.* [Pasifika focus group, P7]

> *The app helps to find a service of your choice if you want to, it's neat.* [Young adult group, P1]

The link to the gambling helpline chat function was highlighted. A participant in the Pasifika focus group commented that "this is like a rescue button. There if you need it." Overall, the ability to access other help (not just the app) was reinforced as a core feature of Manaaki.

Figure 8. Image of the map search feature.

Work Offline

The ability of the app to be accessible offline and not require a Wi-Fi connection at all times was expressed by several participants across all groups:

> *Use will be restricted cause most people do not have data on their phones nor will they spend money [on data] just to use the app. So it is good that you can do stuff that don't need to connect. It's a critical thing, for us.* [Young adult focus group]

Confirming the Final Functions and Features for Manaaki

At the conclusion of the second focus group session, participants confirmed that the core desired features and functions were clearly identified and reflected their insights and experiences. The constant across all groups was that Manaaki would appeal to and be helpful to people seeking help for gambling-related problems. This is illustrated by the following conversation in the Māori focus group:

> *Wow, this is really going to be great for people.* [M1]

> *Yeah, we are awesome aye!* [M2]

> *Yeah, I can't wait to see it all done.* [M1]

> *I am gonna tell people about it, man.* [M3]

> *Ka pai, whānau [well done].* [M4]

Confirming the Final Design Features and Functions Against the Conceptual Framework

The final action undertaken by the research team was to appraise the functions, features, and attributes within the underpinning conceptual framework. Table 2 presents the process outcomes.

Table 2. Confirming the final attributes to be included in Manaaki with associated conceptual frameworks.

Attributes and conceptual framework	Use case examples	Final features and functions for Manaaki
Personalization—developing self-awareness and insight [50-52]	- To be able to add my comments on things that are working well - Put reminders in - To encourage retention, create a longitudinal experience or history as it makes it difficult to stop interacting with the app and difficult to delete - Show where I have come from so it is useful to me - Can add other information that is important to me	- Personal journal that auto-adds from modules plus can add information independently - Shared forum feature that links others in the study (opt-in) - Each module has a progress bar - Videos of people like me within modules and in resources - Use of other languages to help convey complex terms or concepts. Te Reo Māori and Pasifika languages used where relevant - Use of avatars or ability to add own photo
Internal support progress—activating personal strengths and resources and enhancing belief for successful change [54,55]	- To add or interact within the app and receive unanticipated rewards or appreciation that help reinforce behavior - Having information in smaller sections and showing easy actions to promote a sense of capability and then skill development - Messages that pop up that are positive and keep me motivated - Real-time notification messages to reinforce activities and help to interpret or reinforce the activity	- Informative and timely messages and notifications created - Action prompts to complete a topic or module - Instant feedback is given when an activity completed; that is, summary of what the outcomes of an activity are - Assessments can be repeated with scores showing past and current outcomes and an interpretation to help understanding - Able to post specific activities into a journal
External support—relapse prevention [56]	- Can quickly find other help - Resources provided	- Webchat link to national gambling helpline - Additional services section includes links to treatment services - Links to external resources and tools (such as budgeting tools)
Literacy tools—activating personal strengths and resources and enhancing belief for successful change [54,55]	- To have new or novel information presented through videos or notifications or links if needed to read more - Attracting the attention of the user is about relevance and supports the usefulness of the app - Supports how to make changes and maintain these	- Resources and additional help and tools section - Audio options to help when literacy is a challenge (only in English) - Links to other resources such as budgeting and web-based gambling blocker tools - Links to trusted resources - Use of animations - Use of the information icons to provide definitions or interpretation for some words
Navigation—literacy	- Being able to see information in a way that is relevant to me and minimizes unhelpful and inappropriate comparisons - Creates achievements and motivations to keep me engaged - Can add other information that is important to me	- Suggestions of where to next depending on outcomes - A range of distraction activities that are relevant to intended users
Engaging—targeting thoughts and feelings and activating behaviors [53]	- Has a good look and feel - Looks engaging - Tells a story that I can identify with	- Use of New Zealand imagery - Tui guide that provides an explanation of the purpose of each module - Animations
Secure and simple	- It needs to be secure and private - Easy to download - I do not want to be online all the time to use it	- Registration and log-in simple with clear direction - App size low and cater to the lowest operating system as possible - The main elements within the app can be used offline

XSL•FO

RenderX

Discussion

Principal Findings

This paper describes in detail the development of Manaaki, a CBT mobile app intervention designed to support reduction in gambling-related harm. The blending of theory, domain expert input with end-user insights, and lived experience in a collaborative approach is effective for designing mHealth applications [31,36,57]. Creating a visual identity for the new Manaaki app, which resonates with end users, was an important step in establishing and reflecting a sense of trust, connection, and value with the overall intent and purpose of the intervention. To the best of our knowledge, this study is the first to adapt an evidence- and CBT-based program for people experiencing gambling problems into an app-delivered program using a cocreation process with end users.

The demand for mHealth tools that provide interventions (behavior change and cognitive therapy) and generalized support directly to individuals is growing [58-60]. The opportunity to provide interventions for population groups who are unable or unwilling to access more traditional forms is substantial [61,62].

The value of presenting best-practice approaches and evidence in understandable ways (posters, images, and Adobe XD) and cocreating with end users on how they can help translate it into the app's look and feel was an important process for creating a sense of ownership and investment across all aspects of the content, features, function, and format of the final Manaaki app. The use of draft designs and real-time modifiable design flows enabled end-user participants to discuss and identify attributes to be included in the app that would resonate with them. Among the focus group sessions, the research team was able to explore the findings through a theoretical lens, convert the concepts into app designs, and then represent them in ways that were easily understood. Concepts such as personalization were presented using the app user's name and included the creation of a personal avatar. The concept of engagement was reflected in background designs depicting a journey, animation of tools, use of the Tui, and audio features. Empowerment was demonstrated with progress bars, shared forums, self-assessment, and notifications.

Similarly, the content presented back to the participants was adapted and personalized to reflect within-app responses. To support health literacy, words such as fallacy and recovery continued to be used but with the information icon next to them to help participants understand and interpret these terms. In some cases, words and concepts were converted to Reo Māori (the Māori language).

The importance of positively framed and strength-based content was vital. The inclusion of audio files and information icons again elicited positive participant remarks about the app being relevant and appropriate to them. Similar findings are reported elsewhere [63]. The unanticipated prompts (notifications) that support actionable recommendations or app navigation suggestions are tools used in other domains and jurisdictions, such as marketing for supporting retention and motivation [64].

Research on addiction recovery has reported the importance of conveying a sense of social connectedness [65]. The end-user participants in this study also mentioned this, and the inclusion of features such as the webchat, forum, and visibility of services were seen as adding a layer of connection and support within the digital modality.

Adopting learning from the marketing discipline [66] and creating the naming and visual identity for Manaaki (ie, brand) that resonated with participants was essential in establishing and reflecting on them a sense of ownership, trust, connection, and value.

Finally, the attributes of information security, data use, accessibility (ie, available to older mobile phone operating systems), and offline capability were noted as part of the overall app usability discussions; however, they did not overshadow the key aim of the process.

Limitations

Owing to the nature of the research methodology of focus groups, this study also comprises some biased and motivated responses as described in other research using focus groups [67,68]. Nonetheless, our participants were a subset of the target group that focused on the intervention app. The lived experiences of the participants flavored all discussions and decisions and were woven within the final application. Although it may not be possible to suggest that the process and outcome may not adequately represent the wider population of people who experience harm from gambling, it is more likely to resonate than an intervention that does not include any end users [69,70]. Adding a final step in which a small cohort of end users actively used the final app prototype before initiating the clinical trial was not possible in this study because of the overall project timeframe. This step can offer insights that initial design processes cannot anticipate and as such, provide an opportunity to refine the final product by applying active use insights before release, further enhancing the potential for effectiveness.

Conclusions

The importance of domain expert input and evidence-based content combined with end-user insights and involvement should not be undervalued. Weaving these strands and following an iterative process, the research team, with the participants, produced an app that was intentionally tailored to the functional needs of individuals seeking to reduce gambling harm. This process gave the second phase of this study, a randomized waitlist control study to test the effectiveness of Manaaki—a CBT mHealth intervention—the best chance of demonstrating effectiveness. Although adopting this approach does not guarantee success, it is more likely to provide evidence of fidelity to the outcomes of the intervention. Manaaki is currently being trialed using a pragmatic randomized controlled trial with a waitlist control design [43]. This trial was completed in August 2021 and analyses are underway.

Authors' Contributions

GH and CB conceived and wrote the full study grant, which included the cocreation phase and a randomized control trial. JTC, ND, SM, SR, DN, EH, and VN provided expert input for both studies. ND, SM, and SR provided the original GamblingLess content. ND and SM modified the content into 6 modules based on the findings of the GamblingLess study. GH adapted the content to the Manaaki app.

GH designed the cocreation study methodology with inputs from CB, JTC, SEP, and RRC. GH wrote the first draft of this manuscript, and JTC provided the first revisions. All authors reviewed and edited the manuscript. All authors approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Rogers RD, Wardle H, Sharp CA, Wood S, Hughes K, Davies TJ, et al. Gambling as a public health issue in Wales. Bangor, UK: Public Health Wales NHS Trust, Bangor University; 2019.

2. Korn DA, Shaffer HJ. Gambling and the health of the public: adopting a public health perspective. J Gambl Stud 1999;15(4):289-365. [doi: 10.1023/a:1023005115932] [Medline: 12766466]

3. Department of Internal Affairs. Remote interactive gambling and the Gambling Act 2003. Department of Internal Affairs. Wellington: Department of Internal Affairs; 2003. URL: https://www.dia.govt.nz/diawebsite.nsf/wpg_URL/Services-Casino-and-Non-Casino-Gaming-Remote-Interactive-Gambling-and-the-Gambling-Act-2003?OpenDocument [accessed 2021-02-01]

4. Browne M, Bellringer M, Greer N, Kolandai-Matchett K, Rawat V, Langham E, et al. Measuring the burden of gambling harm in New Zealand. Gambling and Addictions Research Centre. Wellington, NZ: Ministry of Health; 2017. URL: https://www.health.govt.nz/system/files/documents/publications/measuring_the_burden_of_gambling_harm_in_new_zealand.pdf [accessed 2021-02-01]

5. Devlin ME, Walton D. The prevalence of problem gambling in New Zealand as measured by the PGSI: adjusting prevalence estimates using meta-analysis. Int Gambl Stud 2012;12(2):177-197. [doi: 10.1080/14459795.2011.653384]

6. Dowling NA, Cowlishaw S, Jackson AC, Merkouris SS, Francis KL, Christensen DR. Prevalence of psychiatric co-morbidity in treatment-seeking problem gamblers: a systematic review and meta-analysis. Aust N Z J Psychiatry 2015;49(6):519-539 [FREE Full text] [doi: 10.1177/0004867415575774] [Medline: 25735959]

7. Lorains FK, Cowlishaw S, Thomas SA. Prevalence of comorbid disorders in problem and pathological gambling: systematic review and meta-analysis of population surveys. Addiction 2011;106(3):490-498. [doi: 10.1111/j.1360-0443.2010.03300.x] [Medline: 21210880]

8. Williams RJ, Volberg RA, Stevens RM. The population prevalence of problem gambling: methodological influences, standardised rates, jurisdictional differences, and worldwide trends. Ontario Problem Gambling Research Centre and the Ontario Ministry of Health and Long Term Care. 2012. URL: https://opus.uleth.ca/bitstream/handle/10133/3068/2012-PREVALENCE-OPGRC%20(2).pdf [accessed 2021-02-01]

9. Thimasarn-Anwar T, Squire H, Trowland H, Martin G. Gambling report: results from the 2016 Health and Lifestyles Survey. Health Promotion Agency. Wellington, NZ: Health Promotion Agency Research and Evaluation Unit; 2017. URL: https://www.hpa.org.nz/sites/default/files/Final-Report_Results-from-2016-Health-And-Lifestyles-Survey_Gambling-Feb2018.pdf [accessed 2021-02-01]

10. Safer gambling aotearoa: me kōrero tātou. Health Promotion Agency. 2021. URL: https://www.safergambling.org.nz/ [accessed 2021-10-18]

11. Suurvali H, Hodgins DC, Toneatto T, Cunningham JA. Hesitation to seek gambling-related treatment among Ontario problem gamblers. J Addict Med 2012;6(1):39-49. [doi: 10.1097/ADM.0b013e3182307dbb] [Medline: 21979819]

12. Abbott M. Gambling and gambling harm in New Zealand: a 28-Year case study. Int J Ment Health Addiction 2017;15(6):1221-1241. [doi: 10.1007/s11469-017-9767-6]

13. Abbott MW, Volberg RA. The New Zealand national survey of problem and pathological gambling. J Gambl Stud 1996;12(2):143-160. [doi: 10.1007/BF01539171] [Medline: 24233913]

14. Alexander BK. Addiction: the urgent need for a paradigm shift. Subst Use Misuse 2012;47(13-14):1475-1482. [doi: 10.3109/10826084.2012.705681] [Medline: 23186447]

15. Morrison L, Boulton A. Reversing the harmful effects of gambling in indigenous families: the development of the Tu Toa Tu Maia intervention. Pimatisiwin 2013;11(2):255-268.

16. Herd R. Tirohanga Taiohi: Taiohi perspectives on gambling among whanua, hapuy, iwi and urban Maori communities. Auckland University of Technology. 2018. URL: https://openrepository.aut.ac.nz/handle/10292/11675 [accessed 2020-11-30]

17. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]

18. mHealth: use of appropriate digital technologies for public health: report by the Director-General. World Health Organization. 2017. URL: https://apps.who.int/iris/handle/10665/274134 [accessed 2020-11-30]

19. Number of mHealth app downloads worldwide from 2013 to 2017 (in billions). Statista. 2019. URL: https://www.statista.com/statistics/625034/mobile-health-app-downloads/ [accessed 2020-05-19]

20. Mobile health (mHealth) market to reach USD 311.98 billion by 2027. Reports and Data. 2020. URL: https://www.reportsanddata.com/report-detail/mobile-health-mhealth-marketReports [accessed 2020-11-09]

21. Unmasking uninstalls: three data points to think about. Adjust. 2018. URL: https://www.adjust.com/blog/unmasking-uninstalls/ [accessed 2020-05-19]

22. Modave F, Bian J, Leavitt T, Bromwell J, Harris Iii C, Vincent H. Low quality of free coaching apps with respect to the American college of sports medicine guidelines: a review of current mobile apps. JMIR Mhealth Uhealth 2015;3(3):e77 [FREE Full text] [doi: 10.2196/mhealth.4669] [Medline: 26209109]

23. Domnich A, Arata L, Amicizia D, Signori A, Patrick B, Stoyanov S, et al. Development and validation of the Italian version of the Mobile Application Rating Scale and its generalisability to apps targeting primary prevention. BMC Med Inform Decis Mak 2016;16:83 [FREE Full text] [doi: 10.1186/s12911-016-0323-2] [Medline: 27387434]

24. Boudreaux ED, Waring ME, Hayes RB, Sadasivam RS, Mullen S, Pagoto S. Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations. Transl Behav Med 2014;4(4):363-371 [FREE Full text] [doi: 10.1007/s13142-014-0293-9] [Medline: 25584085]

25. Savic M, Best D, Rodda S, Lubman DI. Exploring the focus and experiences of smartphone applications for addiction recovery. J Addict Dis 2013;32(3):310-319. [doi: 10.1080/10550887.2013.824331] [Medline: 24074196]

26. Humphreys G, Evans R, Makin H, Cooke R, Jones A. Identification of behavior change techniques from successful web-based interventions targeting alcohol consumption, binge eating, and gambling: systematic review. J Med Internet Res 2021;23(2):e22694 [FREE Full text] [doi: 10.2196/22694] [Medline: 33560243]

27. Michie S, Atkins L, West R. The behaviour change wheel: a guide to designing interventions. Sutton, UK: Silverback Publishing; 2014.

28. Alwashmi MF, Hawboldt J, Davis E, Fetters MD. The iterative convergent design for mobile health usability testing: mixed methods approach. JMIR Mhealth Uhealth 2019;7(4):e11656 [FREE Full text] [doi: 10.2196/11656] [Medline: 31025951]

29. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]

30. Te Morenga L, Pekepo C, Corrigan C, Matoe L, Mules R, Goodwin D, et al. Co-designing an mHealth tool in the New Zealand Māori community with a "Kaupapa Māori" approach. AlterNative 2018;14(1):90-99. [doi: 10.1177/1177180117753169]

31. Eyles H, Jull A, Dobson R, Firestone R, Whittaker R, Te Morenga L, et al. Co-design of mHealth delivered interventions: a systematic review to assess key methods and processes. Curr Nutr Rep 2016;5(3):160-167. [doi: 10.1007/s13668-016-0165-7]

32. Motamarri S. Consumer co-creation of value in mHealth (Mobile Health) service. J Creating Value 2017;3(1):63-76. [doi: 10.1177/2394964317697611]

33. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]

34. Thorsten R, DeFillippi R, Samson A. Managing your co-creation mix: co-creation ventures in distinctive contexts. Eur Bus Rev 2013;25(1):20-41. [doi: 10.1108/09555341311287727]

35. Rabbi M, Philyaw-Kotov M, Li J, Li K, Rothman B, Giragosian L, et al. Translating behavioral theory into technological interventions: case study of an mHealth app to increase self-reporting of substance-use related data. arXiv (forthcoming) 2016:-. [doi: 10.48550/arXiv.2003.13545]

36. Ni Mhurchu C, Te Morenga L, Tupai-Firestone R, Grey J, Jiang Y, Jull A, et al. A co-designed mHealth programme to support healthy lifestyles in Māori and Pasifika peoples in New Zealand (OL@-OR@): a cluster-randomised controlled trial. Lancet Digit Health 2019;1(6):e298-e307 [FREE Full text] [doi: 10.1016/S2589-7500(19)30130-X] [Medline: 33323252]

37.   ECRI Institute-Penn Medicine Evidence-based Practice Center. PCORI evidence map: the impact of mHealth for self-management of chronic disease on patient-centered outcomes. Pcori. 2018. URL: https://www.pcori.org/sites/default/files/PCORI-mHealth-Chronic-Disease-Evidence-Maps-Summary-Report.pdf [accessed 2021-05-30]

38.   Palmer VJ, Weavell W, Callander R, Piper D, Richard L, Maher L, et al. The Participatory Zeitgeist: an explanatory theoretical model of change in an era of coproduction and codesign in healthcare improvement. Med Humanit 2019;45(3):247-257 [FREE Full text] [doi: 10.1136/medhum-2017-011398] [Medline: 29954854]

39.   Verbiest ME, Corrigan C, Dalhousie S, Firestone R, Funaki R, Goodwin D, et al. Using codesign to develop a culturally tailored, behavior change mHealth intervention for indigenous and other priority communities: a case study in New Zealand. Transl Behav Med 2019;9(4):720-736. [doi: 10.1093/tbm/iby093] [Medline: 30388262]

40.   Merkouris SS, Rodda SN, Austin D, Lubman DI, Harvey P, Battersby M, et al. GAMBLINGLESS: FOR LIFE study protocol: a pragmatic randomised trial of an online cognitive-behavioural programme for disordered gambling. BMJ Open 2017;7(2):e014226 [FREE Full text] [doi: 10.1136/bmjopen-2016-014226] [Medline: 28235970]

41.   Dowling N, Merkouris S, Rodda S, Smith D, Lavis T, Austin D, et al. Development and evaluation of an online gambling self-directed program: effective integration into existing services. Victorian Responsible Gambling Foundation. Melbourne, Australia: Deakin University; 2018. URL: https://responsiblegambling.vic.gov.au/documents/348/Development_and_evaluation_of_an_online_gambling_self-direct_help_program.pdf [accessed 2020-05-21]

42.   Dowling NA, Merkouris SS, Rodda SN, Smith D, Aarsman S, Lavis T, et al. Gamblingless: a randomised trial comparing guided and unguided internet-based gambling interventions. J Clin Med 2021;10(11):2224 [FREE Full text] [doi: 10.3390/jcm10112224] [Medline: 34063826]

43.   Humphrey G, Chu J, Dowling N, Rodda S, Merkouris S, Parag V, et al. Manaaki - a cognitive behavioral therapy mobile health app to support people experiencing gambling problems: a randomized control trial protocol. BMC Public Health 2020;20(1):191 [FREE Full text] [doi: 10.1186/s12889-020-8304-x] [Medline: 32028926]

44.   Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983;51(3):390-395. [doi: 10.1037//0022-006x.51.3.390] [Medline: 6863699]

45.   Green W, Cluley R, Gasparin M. Mobile market research applications as a new voice of customer method: implications for innovation and design management. Res Technol Manag 2019;63(1):49-55. [doi: 10.1080/08956308.2020.1686297]

46.   Tong S, Luo X, Xu B. Personalized mobile marketing strategies. J Acad Mark Sci 2020;48(1):64-78. [doi: 10.1007/s11747-019-00693-3]

47.   Ind N, Coates N. The meanings of co-creation. Eur Bus Rev 2013;25(1):86-95. [doi: 10.1108/09555341311287754]

48.   Gorini A, Mazzocco K, Triberti S, Sebri V, Savioni L, Pravettoni G. A P5 approach to m-Health: design suggestions for advanced mobile health technology. Front Psychol 2018;9:2066 [FREE Full text] [doi: 10.3389/fpsyg.2018.02066] [Medline: 30429810]

49.   Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2006;27(2):237-246. [doi: 10.1177/1098214005283748]

50.   Musiat P, Hoffmann L, Schmidt U. Personalised computerised feedback in E-mental health. J Ment Health 2012;21(4):346-354. [doi: 10.3109/09638237.2011.648347] [Medline: 22315961]

51.   DiClemente CC, Marinilli AS, Singh M, Bellino LE. The role of feedback in the process of health behavior change. Am J Health Behav 2001;25(3):217-227. [doi: 10.5993/ajhb.25.3.8] [Medline: 11322620]

52.   Prochaska JO, DiClemente CC, Velicer WF, Rossi JS. Standardized, individualized, interactive, and personalized self-help programs for smoking cessation. Health Psychol 1993;12(5):399-405. [doi: 10.1037//0278-6133.12.5.399] [Medline: 8223364]

53.   Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]

54.   Battersby M, Oakes J, Tolchard B, Forbes A, Pols R. Cognitive behavioral therapy for problem gamblers. In: Zangeneh M, Blaszczynski A, Turner NE, editors. In the pursuit of winning: problem gambling theory, research and treatment. Boston, MA: Springer; 2008:179-197.

55.   Rizeanu S. Cognitive-behavioral therapy for gambling addiction. In: Şenormancı Ö, Şenormancı G, editors. Cognitive behavioral therapy and clinical applications. London, UK: IntechOpen; 2018.

56.   Larimer ME, Palmer RS, Marlatt GA. Relapse prevention. An overview of Marlatt's cognitive-behavioral model. Alcohol Res Health 1999;23(2):151-160 [FREE Full text] [Medline: 10890810]

57.   Dobson R, Herbst P, Candy S, Brott T, Garrett J, Humphrey G, et al. Understanding end-user perspectives of mobile pulmonary rehabilitation (mPR): cross-sectional survey and interviews. JMIR Form Res 2019;3(4):e15466 [FREE Full text] [doi: 10.2196/15466] [Medline: 31859681]

58.   mHealth economics 2017/2018: current status and future trends in mobile health. Research2Guidance. 2017. URL: https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/ [accessed 2020-11-30]

59.   Smith A. Record shares of Americans now own smartphones, have home broadband. Pew Research Center. 2017. URL: https://www.pewresearch.org/fact-tank/2017/01/12/evolution-of-technology [accessed 2020-11-04]

60.   Xu W, Liu Y. mHealthApps: a repository and database of mobile health apps. JMIR Mhealth Uhealth 2015;3(1):e28 [FREE Full text] [doi: 10.2196/mhealth.4026] [Medline: 25786060]

61.   Mobile app of CBT for anxiety and cancer. Clinical Trials. 2014. URL: https://clinicaltrials.gov/show/NCT02286466, [accessed 2021-05-10]

62.   Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019;10(10):CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]

63.   Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. JMIR Mhealth Uhealth 2019;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]

64.   Costacurta RZ. The effect of promotional push notifications from supermarket branded apps on purchase behavior. Universidade de Lisboa. 2019. URL: http://hdl.handle.net/10400.5/19150 [accessed 2020-11-30]

65.   Buckingham SA, Frings D, Albery IP. Group membership and social identity in addiction recovery. Psychol Addict Behav 2013;27(4):1132-1140. [doi: 10.1037/a0032480] [Medline: 23586453]

66.   Batra R, Keller KL. Integrating marketing communications: new findings, new lessons, and new ideas. J Mark 2016;80(6):122-145. [doi: 10.1509/jm.15.0419]

67.   Myers G. Displaying opinions: topics and disagreement in focus groups. Lang Soc 1998;27(1):85-111. [doi: 10.1017/s0047404500019734]

68.   Smithson J. Using and analysing focus groups: limitations and possibilities. Int J Soc Res Methodol 2000;3(2):103-119. [doi: 10.1080/136455700405172]

69.   Still B, Crane K. Fundamentals of user-centered design: a practical approach. Boca Raton, FL: CRC Press; 2017.

70.   Trischler J, Pervan SJ, Kelly SJ, Scott DR. The value of codesign: the effect of customer involvement in service design teams. J Serv Res 2017;21(1):75-100. [doi: 10.1177/1094670517714060]

Abbreviations

CBT: cognitive behavioral therapy
mHealth: mobile health

Original Paper

Patient- and Provider-Reported Experiences of a Mobile Novel Digital Therapeutic in People With Opioid Use Disorder (reSET-O): Feasibility and Acceptability Study

Sarah Kawasaki[1], MD; Sara Mills-Huffnagle[1], MA; Nicole Aydinoglo[2], BA; Halley Maxin[1], BA; Edward Nunes[2], MD

[1]Department of Psychiatry and Behavioral Health, Penn State Health, Hershey, PA, United States
[2]Department of Psychiatry, Columbia University, New York, NY, United States

Corresponding Author:
Sarah Kawasaki, MD
Department of Psychiatry and Behavioral Health
Penn State Health
500 University Drive
Box H073
Hershey, PA, 17033
United States
Phone: 1 7177826844
Fax: 1 7172215674
Email: skawasaki@pennstatehealth.psu.edu

Abstract

Background: Medications for the treatment of opioid use disorder, such as buprenorphine, are effective and essential for addressing the opioid epidemic. However, high dropout rates from medication remain a challenge. Behavioral treatment with contingency management and cognitive behavioral counseling has shown promise for improving the outcomes of buprenorphine treatment but is complicated to deliver. The delivery of behavioral treatment through technology-based platforms has the potential to make it more feasible for widespread dissemination.

Objective: reSET-O is a prescription digital therapeutic and a commercial adaptation of the Therapeutic Education System, an internet-based program with a Community Reinforcement Approach to cognitive behavioral therapy. It delivers cognitive behavioral therapy modules and contingency management rewards upon completion of modules and negative urine drug screens. This pilot study aims to assess the feasibility and acceptability of reSET-O in a community-based opioid treatment program with a Hub and Spoke model of care as part of a larger strategy to maintain individuals in treatment. Objective and qualitative results, as well as acceptability and likeability of reSET-O, were obtained from 15 individuals.

Methods: English-speaking individuals aged ≥18 years with a diagnosis of current opioid use disorder were recruited after being on buprenorphine for at least 1 week of treatment. Two 12-week prescriptions for reSET-O were written for the 24-week study. Patient reports of drug use and likeability scales of reSET-O were conducted at weeks 4, 8, 12, and 24 of the study. Qualitative interviews were also conducted. A total of 4 providers were recruited and provided feedback on the acceptability and feasibility of reSET-O.

Results: Of the 15 participants who participated in this pilot study, 7 (47%) completed 24 weeks, and 8 (53%) were unable to complete because of dropout after enrollment, attrition in treatment, or incarceration. An average of US $96 in contingency management rewards were earned by participants for the completion of modules for the duration of the pilot study. Participants' subjective feedback revealed that reSET-O was easy to use, enjoyable, and helped provide a safe space to admit recurring substance use.

Conclusions: reSET-O was well accepted based on patient and provider feedback in this pilot study; however, adherence and retention in treatment remain areas for improvement. Randomized control trials are needed to assess whether retention of community-based buprenorphine treatment is enhanced through the use of technology-based behavioral interventions such as reSET-O.

XSL•FO
RenderX

Introduction

The opioid epidemic has spread rapidly over the past decade, reaching virtually every region of the United States. Approximately 2.4 million Americans are currently addicted to opioids, and the prognosis is poor; if left untreated, the risk of death from opioid overdose is high [1]. Although different types of medications for opioid use disorder (MOUDs), such as buprenorphine–naloxone, buprenorphine, methadone, and extended-release naltrexone, may be remarkably effective if patients adhere to treatment, adherence to treatment remains challenging [2]. Retention rates in treatment range from 10% to 60%, depending on the medication type [3-14]. In addition to the challenges in adherence, challenges in accessing MOUD present an obstacle not easily surmounted. Methadone can only be dispensed under special regulations that are burdensome to patients (eg, daily attendance at a clinic is initially required). Buprenorphine, available by prescription from any licensed physician who, up until recently, has completed an 8-hour waiver training, has struggled to penetrate primary medical care settings. The supply of physicians, nurse practitioners, and physician assistants prescribing buprenorphine remains limited, and the use of a buprenorphine waiver to the maximum extent of allowed patients is underused [15]. A combination of universally available MOUD along with strategies to combat MOUD treatment attrition and adherence is critical to combat the opioid epidemic.

A key systemic barrier that has been identified is the lack of access to behavioral intervention and counseling to accompany MOUD prescribing [16-18]. Behavioral interventions have the potential to address poor adherence to medication. The provision of counseling is a regulatory requirement for methadone, buprenorphine, and buprenorphine–naloxone treatment. Furthermore, evidence suggests that counseling and behavioral treatments improve the adherence to and effectiveness of MOUD [19], particularly with contingency management (CM) approaches [20-22]. Primary care practices and other clinical settings that are new to addiction treatment typically lack staff with expertise in relevant behavioral treatments, and this gap in care contributes to a reluctance to treat patients with opioid use disorder (OUD) in this setting [16]. Even specialty addiction treatment programs may struggle to deliver more than rudimentary counseling because of time constraints and a lack of expertise in the latest evidence-based interventions. Different models of care have been implemented to address this with success. Vermont established a Hub and Spoke system [23] meant to provide support for practices that may have had barriers to successful outcomes. In this system, the *hub* acts as the specialty treatment center, initiating or escalating OUD medication treatment quickly with the ability to provide the most intense care, including MOUD and therapy, whereas *spoke* sites are primary care practices that may continue patient MOUD after stabilization. Penn State Health partnered with an opioid treatment program (OTP) at the Pennsylvania Psychiatric Institute to use State-Targeted Response funds to establish a Hub and Spoke program [24]. This has allowed for the coordinated care and expansion of MOUDs based in Harrisburg, Pennsylvania, and surrounding counties, including rural regions in south-central Pennsylvania.

Despite the implementation of Hub and Spoke systems, challenges related to behavioral interventions persist, including limited counselor capacities at *hub* sites and potentially no counseling services at *spoke* sites. Mobile apps that deliver behavioral interventions may be beneficial to OUD treatment, helping to fill the gap in the provision of behavioral interventions [25,26]. reSET-O, generated by Pear Therapeutics, Inc, is a comprehensive cognitive behavioral treatment delivered through a mobile phone–based app, with an evidence base suggesting that it can improve the adherence and outcome of MOUD treatment for OUD compared with standard counseling alone [27]. reSET-O is derived from the Therapeutic Education System (TES), developed by academic investigators as a web-based tool delivered by interacting with a computer, delivering a combination of CM—modest monetary or nonmonetary rewards for completion of therapy modules and producing negative drug urine—with cognitive behavioral counseling based on the Community Reinforcement Approach (CRA). reSET-O is a commercial version of TES, adapted for marketing and widespread use, and is delivered as a mobile-based app. reSET-O is available through prescription, and the cost is intended to be covered by insurance. Importantly, the cost of contingent rewards is built into the third-party reimbursement.

CM has been well-established as effective for the treatment of substance use disorders (SUDs) [20,28]; however, funding for incentive rewards has been almost exclusively provided by research grants. The problem of how to fund contingent rewards has stymied the application of CMs in real-world treatment. Thus, the fact that reSET-O can fund incentives by bundling costs into third-party payments is an important advancement. Cognitive behavioral counseling delivered by reSET-O is modeled after the CRA, focusing on cognitive behavioral strategies to achieve abstinence from drugs and build a healthy lifestyle. As approved by the Penn State Health College of Medicine institutional review board, this study piloted the feasibility and acceptability of reSET-O in conjunction with buprenorphine management to (1) assess how individuals in treatment in this Hub and Spoke clinic would interact with this novel intervention and (2) inform future, larger controlled trials using reSET-O.

Methods

Overview

An uncontrolled, 24-week pilot feasibility and acceptability study that added reSET-O to standard care for patients with OUD initiating buprenorphine in an OTP and serving as the hub in a Hub and Spoke system was conducted. The *hub* provides both methadone and buprenorphine treatment in an outpatient clinic specializing in medication treatment for OUDs.

reSET-O was prescribed for a 12-week period and was then renewed for a second 12 weeks for each participant to reach a total prescription duration of 24 weeks. English-speaking individuals aged ≥18 years who were on buprenorphine treatment for at least 1 week were eligible to participate and enroll in the pilot study.

reSET-O is a commercially available prescription digital therapeutic that delivers interactive, self-paced psychoeducational therapeutic modules regarding substance use and is based on the TES. TES is a web-based program that provides written, auditory, and video modules that instill cognitive behavioral skills based on a CRA. The CRA is a therapeutic system that teaches coping skills for staying off drugs and builds skills and activities that are consistent with a healthy lifestyle. The topics of these modules range from managing triggers of substance use and building healthier coping and social skills to HIV prevention and reducing high-risk sexual behavior (see Figure 1 for examples of topics as displayed on the clinician dashboard). Modest monetary rewards (such as gift cards) or nonmonetary rewards are earned for module completion and when negative urine drug screens (USDs) are achieved (see Figure 2 for an example of the reward programming viewed by the participant). With the commercial release of reSET-O, new additions included contingent rewards that may be delivered in the form of gift cards to retail outlets, with the cost covered by third-party insurances. Furthermore, reSET-O prompts daily check-ins, beginning with in-app reports of substance use and cravings, as indicated by the participant (see Figure 3 for an example of endorsed cravings as displayed on the clinician dashboard). Finally, the Patient Services Center, currently referred to as PearConnect, created by Pear Therapeutics, the makers of reSET-O, is a call center that connects patients to a dedicated advocate for support throughout their treatment and prescribers and health care providers to clinician dashboard resources. If a lower activity of reSET-O is detected, Patient Services Center representatives can communicate with patients on a weekly basis to troubleshoot any technical difficulties. reSET-O includes a web-based provider dashboard that permits treatment providers to examine participation and patient engagement remotely (see Figure 4 for an example of a participant snapshot as displayed on the clinician dashboard), as well as facilitate future patient-provider clinic discussions on lessons learned from the modules completed.

Figure 1. reSET-O clinician dashboard and participant lesson progress. This screenshot is an example of the clinician dashboard while viewing an enrolled participant's progress with reSET-O modules. Note that only some of the reSET-O modules are shown in this screen shot. Provided is the title of the module; whether the module was completed, revisited (eg, completed more than once), in progress, or not completed (not pictured); date of completion; and the total amount of time spent on the module. For modules not completed, the Completion Date and Total Time (min) columns would be blank. Copyright 2022 Pear Therapeutics (US), Inc. All rights reserved. Used with permission. Not real patient data.
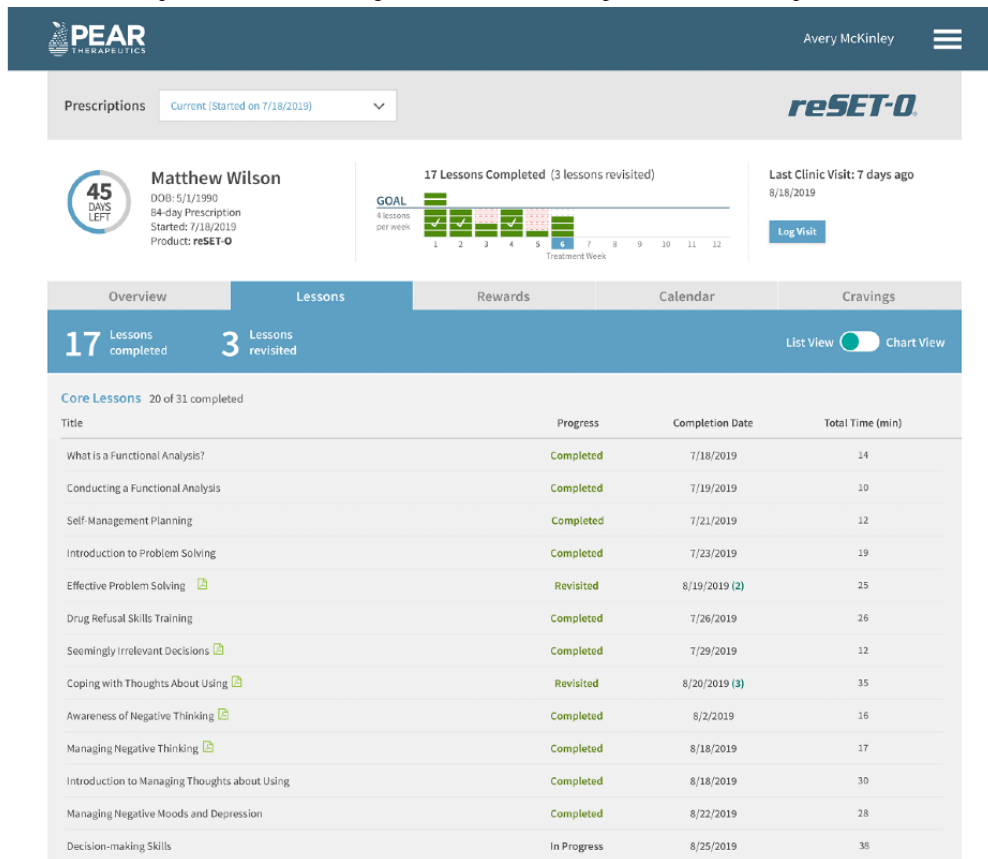
Figure 2. reSET-O participant quiz question and reward screens. This is an example of the type of quiz question that a participant could see (right) while using reSET-O, as well as when the participant was eligible and prompted to spin the wheel (left) to earn a reward of either a monetary amount applied to a specified gift card vendor or written reinforcement. Copyright 2022 Pear Therapeutics (US), Inc. All rights reserved. Used with permission. Not real patient data.
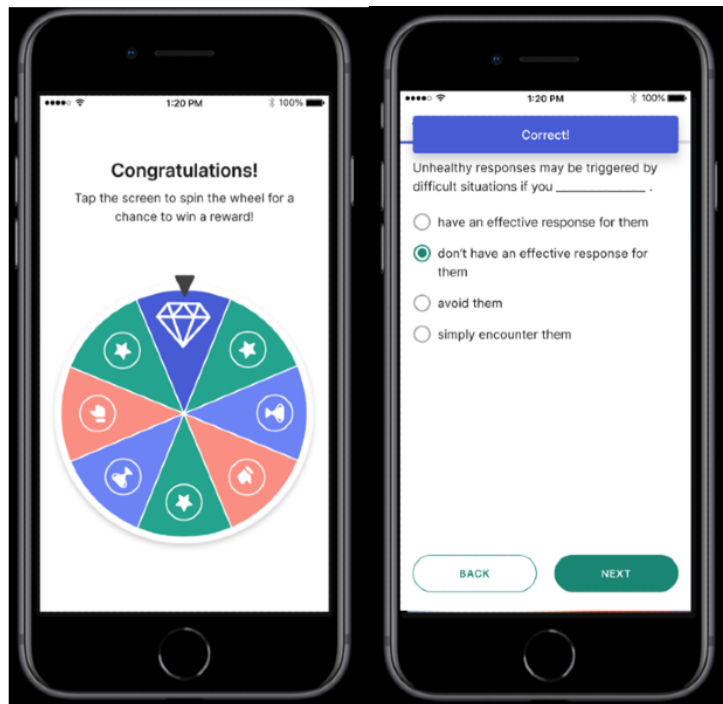
Figure 3. reSET-O clinician dashboard and participant reported cravings. This screenshot is an example of the reSET-O clinician dashboard when viewing a participant's reported cravings while enrolled with reSET-O. After a participant opens reSET-O, they are asked to report any cravings and use, as well as follow-up questions related to craving intensity and potential triggers to use drugs. Copyright 2022 Pear Therapeutics (US), Inc. All rights reserved. Used with permission. Not real patient data.
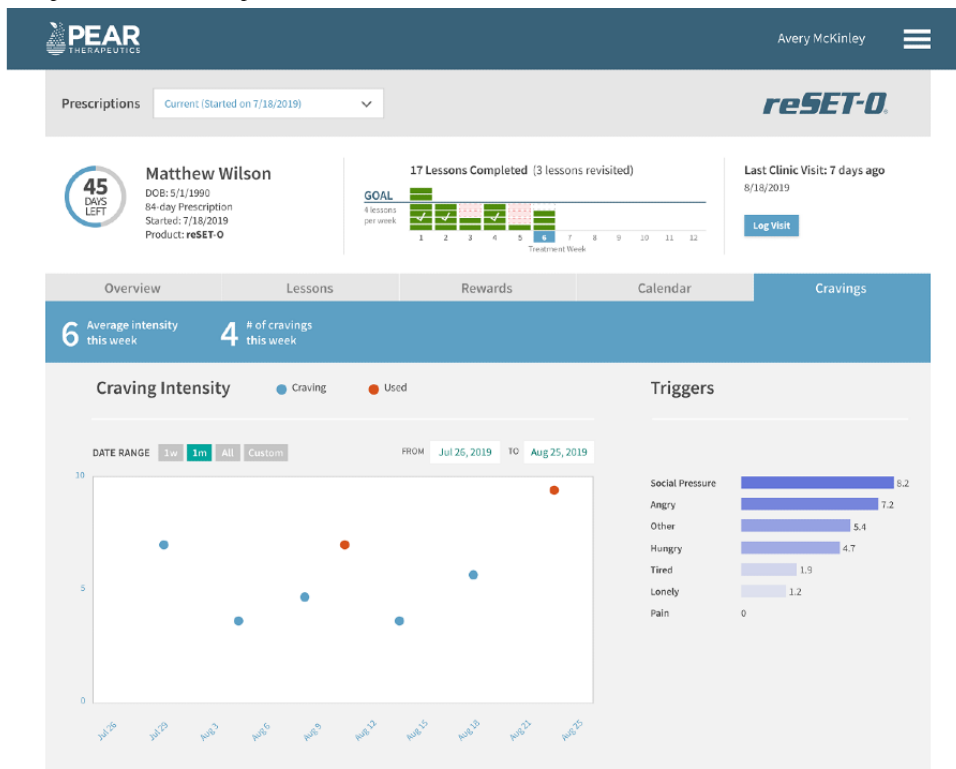
Figure 4. reSET-O clinician dashboard summary and patient snapshot. This is a screenshot of the clinician dashboard while viewing a participant's summary or snapshot data. This information includes days when drug use was self-reported, brief information related to cravings, and the number of lessons completed, stratified by week. Copyright 2022 Pear Therapeutics (US), Inc. All rights reserved. Used with permission. Not real patient data.
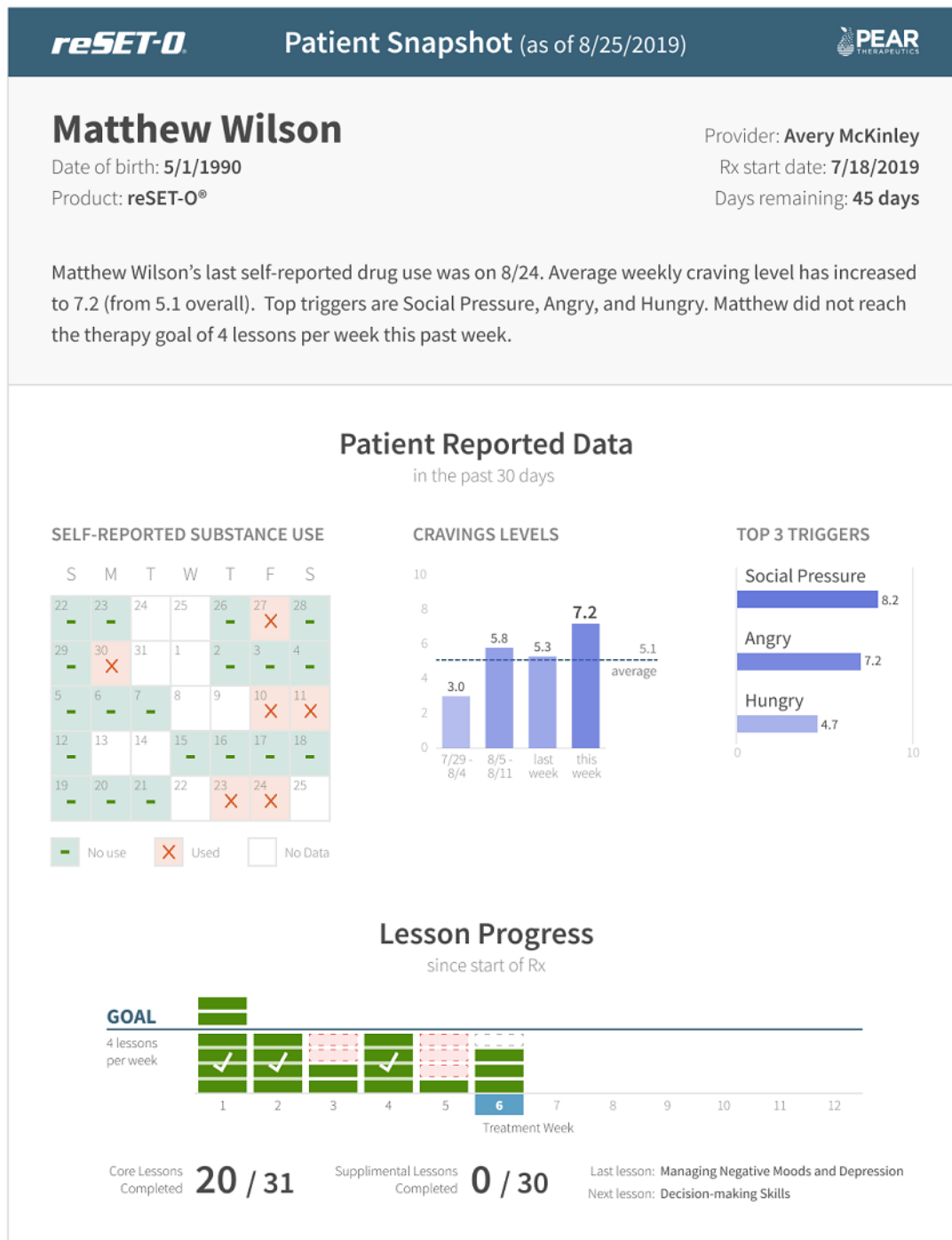
Ethics Approval

Institutional review board approval was obtained for this pilot study from Penn State College of Medicine, under study number 9931 on November 27, 2018, and procedures were followed in accordance with the ethical standards of the institutional committee on human experimentation and with the Helsinki Declaration of 1975, as revised in 2000.

Patient Participants

Eligible individuals were aged ≥18 years; able to read, write, and comprehend English; had a diagnosis of OUD, as determined through routine clinical evaluation; initiated treatment at a community-based OTP serving as the *hub* in a Hub and Spoke system of care; and were prescribed buprenorphine for OUD. We excluded any individuals engaging

in outpatient detoxification or needing a higher level of care, such as inpatient or residential treatment, as well as those not desiring MOUD. Individuals were invited to join the pilot study during their medical appointments and consented in the clinic if interested. All participants provided written informed consent.

Procedures

After obtaining participant consent, reSET-O was prescribed to participants for 12 weeks at a time, with a total of 2 prescriptions (24 weeks of consecutive treatment). reSET-O asks participants to log onto the app and complete 4 learning modules per week, each focusing on a particular skill. The modules are presented in a fixed order, beginning with modules on skills to avoid drugs, followed by modules on building a healthy lifestyle and HIV risk reduction. Each completed module yields a chance to earn rewards in the form of gift cards to retail outlets. Participants were evaluated via the self-report of drug use with the Timeline Follow-Back (TLFB) method, USDs, and three mood assessments: the Kessler-10 [29,30], the Posttraumatic Stress Disorder (PTSD) Checklist–Civilian [31-33], and the Patient Health Questionnaire-9 [34]. These evaluations occurred at baseline and at 4, 8, 12, and 24 weeks after study entry. Self-reported drug use was collected using TLFB [35,36] and verified using USDs. Participants were scored as not abstinent at a visit if either a self-report or a USD was positive for opioids. Within the reSET-O, craving assessments were performed throughout the duration of the prescription.

Participant feedback data regarding reSET-O was collected on the Intervention Acceptability Feedback Form (IAFF), evaluating seven characteristics—interest, usefulness, new information learned, ease of understanding, relevancy, satisfaction, and likeability—on a Likert scale of 0 to 9 (0=lowest and 9=highest). The IAFF was completed at each visit after baseline (weeks 4, 8, 12, and 24).

Qualitative interviews with participants were conducted by research assistants between weeks 8 and 12 to capture information from participants about their experience using the app, acceptability of the app, and suggestions for improvement. Approximately 60% (9/15) of the participants completed the interviews. Each participant was asked 9 standard questions (Textbox 1); then, the staff member could ask follow-up and probing questions as needed to elicit additional information. The interviews lasted between 30 and 45 minutes, were digitally audio recorded, and transcribed verbatim for analysis purposes.

A total of 3 research staff members reviewed all 9 interviews, extracting transcribed text relevant and responsive to the domains of satisfaction or dissatisfaction and the likeability of the intervention. Responses to the following research question: "How did participants like using reSET-O?" were categorized into unique themes, which were reviewed and discussed by a larger research team to reach a consensus.

Textbox 1. Qualitative interview questions and examples of follow-up questions.

Tell me how you use technology in your life.

- What other apps do you use regularly?
- How often do you use other apps?
- Where do you get information about health and/or medical questions?

Think about the times you used the reSET-O app.

- At what times during the day did you use the app? How often?
- When did you access the reSET-O app the most?
- Where did you access the reSET-O app?
- How often did you check in with the app between modules?

What interested you about this study?

- Why did you decide to participate in the study?
- How did it fit in with your counseling at the clinic?
- Did you talk about the reSET-O app with other people (friends, family, or other clinic patients)?

How did you like using the reSET-O app?

- Can you give me an example of what you liked about it?
- What did you not like?
- Would you recommend this app to someone? Why or why not?

How useful/relevant to your life was the reSET-O app?

- Which modules were most useful and/or relevant?
- Which were the least useful and/or relevant? Why did those not work for you?
- Did you repeat any of the modules? Which ones? Why?
- Which features of the app were most useful and/or relevant?
- Which features were the least useful and/or relevant? Why did those not work for you?

Is there anything about the reSET-O app that you would change?

- Content?
- Language?
- Videos?
- Examples?
- What type of changes would you make?

What else would you like the reSET-O app to do?

- Give examples of additional features or module topics not currently available.

Were you able to complete all the study tasks up to this point?

- Was there anyone or anything that helped you complete all of the research study tasks up to this point?
- If not, why were you not able to complete all of the study tasks up to this point?
- Was there anyone or anything that helped you complete some of the research study tasks up to this point?
- Was there anything about the research study that kept you from completing all the study tasks?
- Was there anything about using the reSET-O app that made you not want to continue in the study?
- What was the main reason for you to continue in the study?
- Was there anything else about the research study that made you not want to continue?

Were you able to complete all or some of the reSET-O learning modules up to this point?

- Was there anyone or anything that helped you complete all of the learning modules up to this point?

- If not, why were you able to complete all of the learning modules up to this point?

- Was there anyone or anything that helped you complete some of the learning modules up to this point?

- Was there anything about the reSET-O app that kept you from completing all of the learning modules?

- Was there anything about using the reSET-O app that made you not want to continue the learning modules?

- What was the main reason for you to continue using the app (if applicable)?

- Was there anything else about the research study or the app that kept you from completing the reSET-O modules?

Provider Participants

Provider data regarding the acceptability and feasibility of reSET-O were also collected. Clinicians working in the treatment programs were asked to create their own reSET-O accounts and review the learning modules. After 3 weeks of reSET-O use, data from 4 clinical providers were collected through the Weiner Intervention Acceptability, Appropriateness, and Feasibility form (WIAAF; Multimedia Appendix 1), evaluating the acceptability, appropriateness, and feasibility of the app intervention. Each category was assessed through 4 prompts, each using a 5-point Likert scale (1=lowest and 5=highest). The highest score for any given category could be 20. The WIAAF, currently in development, has no cutoff scores for interpretation; however, higher scores indicate greater acceptability, appropriateness, or feasibility.

Results

Overview

Of the 15 participants, 3 (20%) female patient participants and 12 (80%) male patient participants with OUD were enrolled in the study, with an average age of 36.2 (SD 9.3) years (Table 1). Approximately 73% (11/15) of the participants identified as White, whereas 7% (1/15), 13% (2/15), and 7% (1/15) identified as Black or African American, biracial, or other, respectively, and 13% (2/15) identified as Hispanic (Table 1). Of these 15 participants, 2 (13%) withdrew consent after baseline; 3 (20%) participants dropped out of clinical care before week 8 and were unable to be reached, and another 3 (20%) became incarcerated between weeks 12 and 24 during the study. Thus, of the 15 participants, 8 (53%) withdrew and/or were unable to complete the study, and 7 (47%) were able to complete all 24 weeks of the study.

Table 1. Study demographics (N=15).

Characteristics	Baseline (n=15)	Week 4 (n=11)	Week 8[a] (n=10)	Week 24 (n=7)
Age (years), mean (SD)	36.2 (9.3)	37.4 (9.4)	37 (9.1)	41.4 (8.1)
Gender, n (%)				
Male	12 (80)	11 (100)	10 (100)	7 (100)
Female	3 (20)	0 (0)	0 (0)	0 (0)
Race, n (%)				
White	11 (73)	8 (73)	7 (70)	5 (71)
Black or African American	1 (7)	1 (9)	1 (10)	1 (14)
Biracial	2 (13)	2 (18)	2 (20)	1 (14)
Other[b]	1 (7)	0 (0)	0 (0)	0 (0)
Ethnicity, n (%)				
Hispanic	2 (13)	1 (9)	1 (10)	14

[a]At weeks 8 and 12, the same 10 participants remained in the study.

[b]Participants identified as Hispanic only.

Patient Participant Feedback and Outcomes

Clinical Outcomes

Although all 15 participants had current opioid use at the time of entry into the clinic, at the point of entry into the study, 8 (53%) had already achieved abstinence from opioids through outside treatment, whereas 7 (47%) participants reported and tested positive for opioid use. Of the 7 participants who reported and/or tested positive for opioid use at baseline, 1 (14%) endorsed opiate use at weeks 4, 8, and 24, and 3 (43%) endorsed opioid use at week 12 (Table 2). Of the 8 participants who did not report or test positive for opioid use at baseline, none endorsed opioid use at weeks 4, 8, and 24, whereas 1 (13%) endorsed opioid use at week 12 (Table 2). Of the 15 participants, 7 (47%) participants completed the TLFB at all time points; however, 1 (7%) participant had trouble urinating at some

appointments because of other medical conditions. In this case, only TLFB was used to assess abstinence.

Current symptoms of emotional distress were measured using the self-report screening instruments for depression (Patient Health Questionnaire-9), general emotional distress (Kessler-10), and PTSD symptoms (PTSD Checklist–Civilian were collected at all 5 time points. Means and SDs were calculated and are reported in Table 3. Compared with the baseline, the mean scores decreased over time across all measures, although the scores reflected, at most, mild levels of severity from baseline through the follow-ups.

In application, patient-reported and -endorsed cravings and craving triggers were collected throughout the duration of the reSET-O prescriptions and reported as frequency counts (Table 4). Approximately 53% (8/15) of patients reported cravings within reSET-O. Overall, by the end of the trial, there were fewer reports of endorsed cravings and craving triggers when compared with the beginning of the trial; however, 13% (2/15) of participants continued to report cravings and subsequent triggers throughout both of the reSET-O prescriptions.

Table 2. Opioid positive and negative results based on Timeline Follow-Back and urine drug screens (N=15)[a].

Opioid status at baseline	Baseline (n=15)	Week 4 (n=11)	Week 8 (n=10)	Week 12 (n=10)	Week 24 (n=7)
Participants testing *positive* at baseline (positive for opioids), n/N (%)	7/7 (100)	1/4 (25)	1/4 (25)	3/4 (75)	1/1 (100)
Participants testing *negative* at baseline (positive for opioids), n/N (%)	0/8 (0)	0/7 (0)	0/6 (0)	1/6 (17)	0/6 (0)

[a]Patients were scored as negative if both self-report for the past 30 days by Timeline Follow-Back was negative for opioids and urine toxicology was negative for opioids and scored as positive otherwise.

Table 3. Negative mood scores.

Time point	Participants, n (%)	PHQ-9[a,b], mean (SD)	Kessler-10[c], mean (SD)	PCL-C[d,e], mean (SD)
Baseline	15 (100)	10.33 (6.93)	21.66 (10.76)	17 (14.73)
Week 4	11 (73)	6.46 (6.19)	15.86 (11.03)	14.73 (13.23)
Week 8	10 (67)	5.06 (5.49)	14.66 (11.27)	11.86 (10.84)
Week 12	10 (67)	5.93 (5.79)	14 (11.55)	12.6 (11.87)
Week 24	7 (47)	4.53 (5.57)	10.86 (12.22)	11.6 (14.20)

[a]PHQ-9: The Patient Health Questionnaire-9.

[b]Scores range from 0 (minimal) to –27 (severe) depression asking, "over the last 2 weeks, how often have you been bothered by any of the following problems?" Scores in the 5 to 9 range are considered mild severity.

[c]The Kessler-10 Psychological Distress Scale asks participants how they have been feeling over the past 1 month. Scores range from 10 (minimal) to 50 (severe distress).

[d]PCL-C: Posttraumatic Stress Disorder Checklist–Civilian.

[e]The PCL-C asks, "how much have you been bothered by each of the following 20 statements in the past 1 month?" Scores range from 0 to 80, with scores >30 being likely to have a diagnosis of posttraumatic stress disorder.

Table 4. In-app patient-reported triggers to endorsed cravings (n=8 respondents)[a].

Category	Pain	Anger	Other	Tired	Hungry	Lonely	Social pressure
Patients reporting triggers (first 12 weeks), n (%)	3 (38)	3 (38)	5 (63)	5 (63)	2 (25)	6 (75)	0 (0)
Patients reporting triggers (second 12 weeks), n (%)	1 (13)	0 (0)	1 (13)	2 (25)	0 (0)	1 (13)	0 (0)

[a]Each reSET-O check-in featured craving assessment and triggers that induced cravings to use drugs: pain, anger, fatigue, hunger, isolation, and social pressure.

reSET-O Patient Feedback

After the first 12 weeks, 73% (11/15) of participants had accessed reSET-O, with an average lesson completion of 15.7 lessons, ranging from 0 to 36 completed lessons. The rewards earned during this time averaged US $50, ranging from US $5 to US $120. Approximately 33% (5/15) of participants engaged with reSET-O throughout the 24 weeks of the trial. Among these participants, there was an average lesson completion of

15.2 lessons, ranging from 0 to 37 completed lessons, and an average of US $49 rewards earned, ranging from US $0 to US $85.

It should be noted that individuals who completed all 24 weeks of the study completed, on average, 6.2 lessons per week and repeated modules. Those who dropped out before 12 weeks completed an average of 2.5 lessons per week.

XSL•FO

RenderX

Likeability scores, as assessed by the IAFF, are shown in Figure 5. The IAFF comprises 7 categories, including likeability, which was assessed using a Likert scale ranging from 0 to 9. Anchors were generated based on the prompt. For example, for the likeability category, a score of 0 indicated *I do not like it at all*, and a score of 9 indicated *I like it a lot*. With regard to likeability, after 4 weeks of reSET-O use, the average scores ranged between 6 and 8 out of a maximum possible score of 9 across all categories. The overall trend of scores appeared to drop slightly by the end of the 24 weeks, with an average range of 4.9 to 7.1 across all categories (Figure 5).

Qualitative feedback indicated overall reSET-O feasibility and acceptability by participants, as well as suggestions for improvement. The first identified theme was motivation and attitude regarding treatment delivery (eg, delivery similar to a provider appointment). Participant 3 illuminated this theme stating the following:

> *I also like that the app [reSET-O®] is more personal, I feel like I can be open and honest because it's a person but not a person. I feel like I can say things or endorse things in the app [reSET-O®] that I normally wouldn't say or want to talk about in a group setting* [participant 3, male, 45 years]

The second identified theme was relevancy (eg, content relevant to personal use or life). Participant 9 expressed the following:

> *time management and anxiety. I think there was one on depression and anxiety stuff like that, relates to me* [participant 9, male, 30 years]

The third theme was app features (eg, ease of use). Participant 11 stated the following:

> *the general app [reSET-O®], you know, the bones of it are good, you know. It takes you where you need to be, and it's very easy to use.* [participant 11, male, 40 years]

The last theme was the impact on knowledge, skills, and behavior (eg, module content). Participant 8 expressed the following:

> *People that actually want to stay clean, it [reSET-O®] actually gives you different information that you never knew.* [participant 8, male, 30 years]

Many participants reported liking the rewards earned, the novelty of the treatment, and its similarity to provider treatments. Some participants found the content relevant to their own situations, and most felt that reSET-O was very easy to use. Additional analyses demonstrated that, overall, reSET-O felt personal to participants with relevant content and ease of use.

Areas of improvement were also suggested by participants during the qualitative interviews. Reported problematic areas with reSET-O included varying individual relatability to any module, as noted by a participant that the modules sometimes "felt like a drag," and a burdensomely persistent notification system if a participant did not use the app by an expected time point.

Figure 5. Average participant Intervention Acceptability Feedback Form scores. Participants were asked to rate the reSET-O in each of the 7 categories listed on the x-axis. Overall, the scores fluctuated in the range of 5 to 9 out of a maximum possible score of 9, suggesting moderately good acceptability on all dimensions. Scores appeared to diminish over time, especially during week 24.
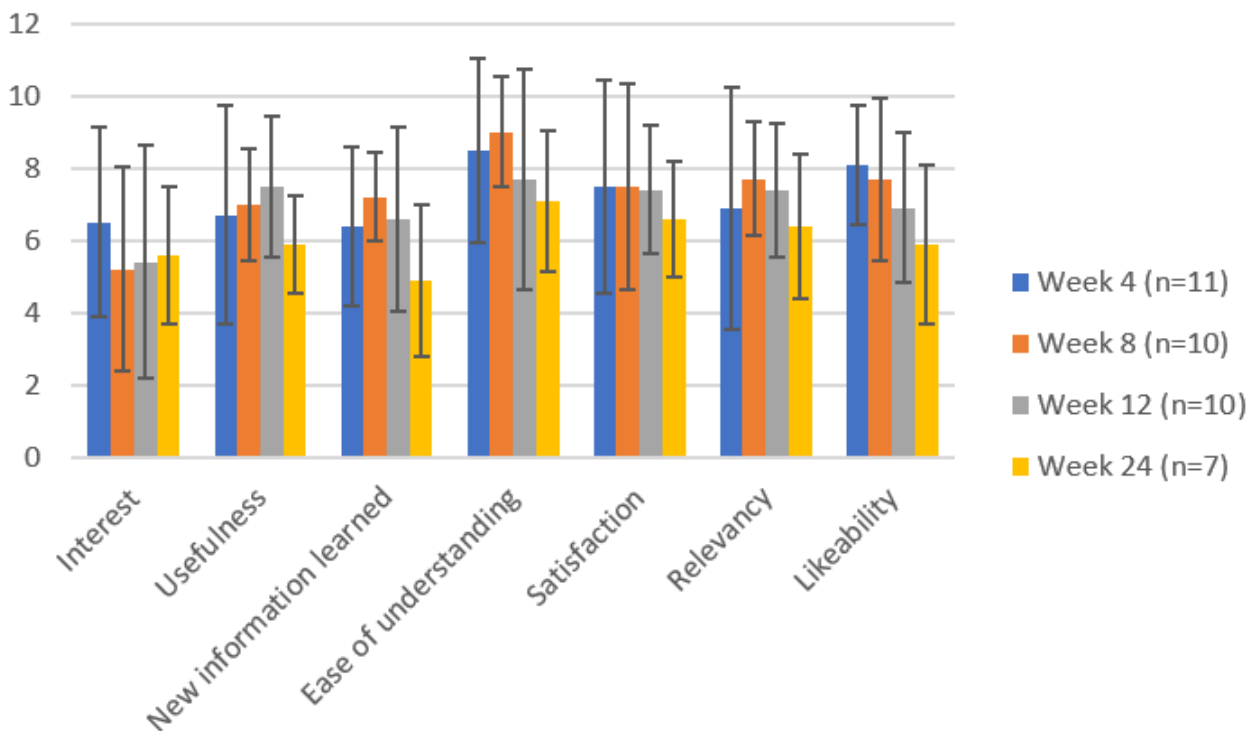
reSET-O Provider Feedback

Provider feedback was collected on the WIAAF after 3 weeks of reSET-O use to gauge provider insights on the acceptability, appropriateness, and feasibility of reSET-O; the average score for the acceptability of reSET-O was 13.8 (SD 2.1), whereas appropriateness and feasibility were 12.5 (SD 2.5) and 15.8 (SD 0.5), respectively. Each category was assessed by 4 prompts, which were measured using a Likert scale ranging from 1 (completely disagree) to 5 (completely agree), and these scores were summed for a maximum score of 20. Although the WIAAF does not have established or validated cutoff scores for interpretation, the higher the score, the greater the acceptability, appropriateness, and feasibility. These scores suggest acceptability and feasibility in the good range but are short of the maximum.

Qualitative feedback from providers suggested that providers liked reSET-O and saw its potential as being implemented in a clinical setting. However, areas for improvement were also noted. Some providers indicated that the content and examples felt as though they were written by someone who had not experienced substance use and recovery. For example, providers expressed that some of the language used was not at an appropriate reading level and that the language used was stigmatizing. For instance, the phrase *beating drug addiction* was reported as pejorative, as if substance use was simply something to overcome and move on with one's life as opposed to a chronic illness, for which "life changes and long-term commitment [were] needed to sustain recovery," as quoted by an individual. Providers also indicated that it would be useful to have an alert option within reSET-O to alert a provider or recovery support person when a patient is in an emergency situation, in distress, or in need of support. It was also suggested that when starting a new module, a review of the last completed module should occur before beginning the new module content.

Discussion

Principal Findings

In this preliminary feasibility pilot study among patients undergoing buprenorphine treatment for OUD, reSET-O appeared generally acceptable and feasible, warranting further study. However, there was still a high rate of dropout, and the extent to which patients used the app varied. Some patients used the learning modules and appreciated the cognitive behavioral skills covered, whereas others did not engage. Opioid use was low among patients who remained in the study, although many of these patients had already achieved abstinence at the study baseline. Depression and anxiety symptoms were within the mild range of severity, with trends toward improvement. In addition, across all 24 weeks, the participant IAFF satisfaction ratings were between 5 and 9, out of a total of 9, for the seven assessed areas (interest, usefulness, new information learned, ease of understanding, satisfaction, relevance, and likeability), suggesting a good level of satisfaction. These scores decreased slightly over time for all assessed categories, which may be a reflection of the decay in enthusiasm over time. There was a high rate of dropout from the study, with 91% (10/11) of the patients who initiated reSET-O completing 12 weeks of

reSET-O and 64% (7/11) completing 24 weeks. This is perhaps not surprising, given the general problem of attrition from apps, as well as the high rates of dropout from the treatment characteristics of patients with SUDs. However, this suggests that strategies are needed to improve engagement and retention in treatment with reSET-O.

One such strategy could be more integration between counseling sessions with clinical providers and the therapy modules delivered by the app. reSET-O is largely psychoeducational and didactic in nature. Future studies should invite counselors and medical providers to discuss reSET-O and the topics therein with patients to make the content more relevant and engaging. Clinicians prescribing medication routinely discuss medication adherence, side effects, and whether and how it may be helpful. Similarly, clinicians prescribing reSET-O could ask about what modules the patient has completed, troubleshoot adherence problems, and discuss how the patient is applying the therapy modules delivered by the app in their daily lives. In this way, in-person and mobile treatments can enhance one another to perhaps increase treatment retention. Some patients suggested additional praise or encouragement for reaching certain milestones, such as maintaining a certain number of favorable urines or reaching a specified length of time of sobriety, which could be further facilitated by counselors. A team approach among providers, patients, and reSET-O may better address any problems with adherence, perceived burden, and the relevance of the topics. The integration of remote therapy with in-person treatment might enhance the salience of reSET-O and reduce enthusiasm decay over time.

Second, provider feedback, although encouraging, provided some helpful suggestions on language. On the basis of the WIAAF scores, the feasibility of implementing reSET-O had the highest average score when compared with the other categories. The category with the lowest average score was appropriateness because of the content language. As noted above, some providers felt that the language appeared stigmatizing in some areas or was difficult to understand (eg, reading level). TES was created >15 years ago, and much has changed in the culture and lexicon surrounding SUDs. This presents a challenge in terms of how a digital therapeutic could be adjusted based on the patient population and setting, perhaps based on real-time feedback from local patients and providers, to provide a product that is maximally inviting and relevant. Regarding the scope of services, providers felt that emergency information might be beneficial to the patient. Mobile technology in this arena has been developed through the Addiction–Comprehensive Health Enhancement Support System [28], and perhaps, interapplication communication can broaden the scope to help more patients.

The monetary remuneration for participants was relatively low compared with traditional evidence-based CM programs [20,37]. reSET-O rewards negative USDs and module completions, with the idea that incentives will help patients participate in the therapy modules; thus, internalizing techniques of cognitive behavioral therapy can prevent ongoing drug use and address problems with cravings, mood, and relationships. Traditionally, CM has shown efficacy in rewarding individuals for appointment adherence and negative USDs alone. Moreover,

larger rewards were available in prior grant-funded studies [38,39]. That reSET-O funds incentives through third-party reimbursement is an innovation that overcomes the problem of how to fund incentives and makes CM feasible in community-based practice. However, only modest reward values are possible using this approach. reSET-O, and its forerunner TES, use the *prize bowl* model for low-cost CM developed by Petry [37]. This turns CM into a game where rewards are determined by chance; some are just verbal (eg, *good job*), and most rewards have low monetary value, approximately US $5, with an occasional larger reward value, which keeps the overall costs of contingencies low across treatment episodes.

Limitations

Limitations include that this study was an uncontrolled trial with a small sample size and a high attrition rate. This study is most useful for generating ideas for improving feasibility and acceptability; however, without a control group, it is not possible to evaluate effectiveness. Patients retained in the study were mostly abstinent over the course of the study, and mood ratings were in the mild range of severity, which is encouraging. However, it is not possible to evaluate treatment retention or patients' individual trends in substance use and mood without a larger sample and control condition. The fact that many patients who were abstinent during the study were already abstinent at baseline suggests the importance of covarying for

baseline in future trials. The sample comprised predominantly White men, reflective of the local patient population, affiliated with the one particular clinic in which the study was conducted. The providers did not systematically discuss reSET-O with the enrolled patients during clinical visits. Participants might achieve more therapeutic benefits from reSET-O if it is integrated more into in-person counseling.

Conclusions

Prescription digital therapeutics, such as reSET-O, have the potential to bridge a gap in MOUD, which often prevents potential providers from prescribing these life-saving medications, namely by addressing the need to deliver medication in conjunction with behavioral counseling. For providers who practice without live behavioral counseling on site and within financially stressed clinical programs, a mobile app has the potential to expand a clinic's ability to meet the needs of populations with SUDs. Larger controlled trials are warranted to evaluate whether this intervention improves the adherence to and outcome of MOUD in community-based treatment settings such as the Hub and Spoke model of care in which this pilot study took place. Future work should examine ways that clinicians can integrate the patients' participation in an app such as reSET-O into the counseling of their patients in an effort to improve adherence to the use of the app and maximize its impact as a clinician extender.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Wiener Intervention Acceptability, Appropriateness, and Feasibility form.
[DOCX File , 17 KB - formative_v6i3e33073_app1.docx]

References

1. Hser YI, Hoffman V, Grella CE, Anglin MD. A 33-year follow-up of narcotics addicts. Arch Gen Psychiatry 2001;58(5):503-508. [doi: 10.1001/archpsyc.58.5.503] [Medline: 11343531]
2. Veilleux JC, Colvin PJ, Anderson J, York C, Heinz AJ. A review of opioid dependence treatment: pharmacological and psychosocial interventions to treat opioid addiction. Clin Psychol Rev 2010;30(2):155-166. [doi: 10.1016/j.cpr.2009.10.006] [Medline: 19926374]
3. Mattick RP, Breen C, Kimber J, Davoli M. Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence. Cochrane Database Syst Rev 2014(2):CD002207. [doi: 10.1002/14651858.CD002207.pub4] [Medline: 24500948]
4. Hser YI, Saxon AJ, Huang D, Hasson A, Thomas C, Hillhouse M, et al. Treatment retention among patients randomized to buprenorphine/naloxone compared to methadone in a multi-site trial. Addiction 2014;109(1):79-87 [FREE Full text] [doi: 10.1111/add.12333] [Medline: 23961726]
5. Lee JD, Nunes Jr EV, Novo P, Bachrach K, Bailey GL, Bhatt S, et al. Comparative effectiveness of extended-release naltrexone versus buprenorphine-naloxone for opioid relapse prevention (X:BOT): a multicentre, open-label, randomised controlled trial. Lancet 2018;391(10118):309-318 [FREE Full text] [doi: 10.1016/S0140-6736(17)32812-X] [Medline: 29150198]

6. Wakeman SE, Larochelle MR, Ameli O, Chaisson CE, McPheeters JT, Crown WH, et al. Comparative effectiveness of different treatment pathways for opioid use disorder. JAMA Netw Open 2020;3(2):e1920622 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.20622] [Medline: 32022884]

7. Grimm CA. Geographic disparities affect access to buprenorphine services for opioid use disorder. U.S. Department of Health and Human Services, Office of the Inspector General. 2020. URL: https://oig.hhs.gov/oei/reports/oei-12-17-00240.pdf [accessed 2022-03-14]

8. Morgan JR, Schackman BR, Weinstein ZM, Walley AY, Linas BP. Overdose following initiation of naltrexone and buprenorphine medication treatment for opioid use disorder in a United States commercially insured cohort. Drug Alcohol Depend 2019;200:34-39 [FREE Full text] [doi: 10.1016/j.drugalcdep.2019.02.031] [Medline: 31082666]

9. Morgan JR, Schackman BR, Leff JA, Linas BP, Walley AY. Injectable naltrexone, oral naltrexone, and buprenorphine utilization and discontinuation among individuals treated for opioid use disorder in a United States commercially insured population. J Subst Abuse Treat 2018;85:90-96 [FREE Full text] [doi: 10.1016/j.jsat.2017.07.001] [Medline: 28733097]

10. Weiss RD, Potter JS, Fiellin DA, Byrne M, Connery HS, Dickinson W, et al. Adjunctive counseling during brief and extended buprenorphine-naloxone treatment for prescription opioid dependence: a 2-phase randomized controlled trial. Arch Gen Psychiatry 2011;68(12):1238-1246 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2011.121] [Medline: 22065255]

11. Weiss RD, Potter JS, Griffin ML, Provost SE, Fitzmaurice GM, McDermott KA, et al. Long-term outcomes from the national drug abuse treatment clinical trials network prescription opioid addiction treatment study. Drug Alcohol Depend 2015;150:112-119 [FREE Full text] [doi: 10.1016/j.drugalcdep.2015.02.030] [Medline: 25818060]

12. Hser YI, Evans E, Huang D, Weiss R, Saxon A, Carroll KM, et al. Long-term outcomes after randomization to buprenorphine/naloxone versus methadone in a multi-site trial. Addiction 2016;111(4):695-705 [FREE Full text] [doi: 10.1111/add.13238] [Medline: 26599131]

13. Hser YI, Huang D, Saxon AJ, Woody G, Moskowitz AL, Matthews AG, et al. Distinctive trajectories of opioid use over an extended follow-up of patients in a multisite trial on buprenorphine + naloxone and methadone. J Addict Med 2017;11(1):63-69 [FREE Full text] [doi: 10.1097/ADM.0000000000000274] [Medline: 27898496]

14. Larochelle MR, Bernson D, Land T, Stopka TJ, Wang N, Xuan Z, et al. Medication for opioid use disorder after nonfatal opioid overdose and association with mortality: a cohort study. Ann Intern Med 2018;169(3):137-145 [FREE Full text] [doi: 10.7326/M17-3107] [Medline: 29913516]

15. Stein BD, Sorbero M, Dick AW, Pacula RL, Burns RM, Gordon AJ. Physician capacity to treat opioid use disorder with buprenorphine-assisted treatment. JAMA 2016;316(11):1211-1212 [FREE Full text] [doi: 10.1001/jama.2016.10542] [Medline: 27654608]

16. Quest TL, Merrill JO, Roll J, Saxon AJ, Rosenblatt RA. Buprenorphine therapy for opioid addiction in rural Washington: the experience of the early adopters. J Opioid Manag 2012;8(1):29-38 [FREE Full text] [doi: 10.5055/jom.2012.0093] [Medline: 22479882]

17. Hutchinson E, Catlin M, Andrilla CH, Baldwin LM, Rosenblatt RA. Barriers to primary care physicians prescribing buprenorphine. Ann Fam Med 2014;12(2):128-133 [FREE Full text] [doi: 10.1370/afm.1595] [Medline: 24615308]

18. DeFlavio JR, Rolin SA, Nordstrom BR, Kazal Jr LA. Analysis of barriers to adoption of buprenorphine maintenance therapy by family physicians. Rural Remote Health 2015;15:3019 [FREE Full text] [Medline: 25651434]

19. McLellan AT, Arndt IO, Metzger DS, Woody GE, O'Brien CP. The effects of psychosocial services in substance abuse treatment. JAMA 1993;269(15):1953-1959. [Medline: 8385230]

20. Peirce JM, Petry NM, Stitzer ML, Blaine J, Kellogg S, Satterfield F, et al. Effects of lower-cost incentives on stimulant abstinence in methadone maintenance treatment: a national drug abuse treatment clinical trials network study. Arch Gen Psychiatry 2006;63(2):201-208. [doi: 10.1001/archpsyc.63.2.201] [Medline: 16461864]

21. Nunes EV, Ball S, Booth R, Brigham G, Calsyn DA, Carroll K, et al. Multisite effectiveness trials of treatments for substance abuse and co-occurring problems: have we chosen the best designs? J Subst Abuse Treat 2010;38 Suppl 1:S97-112 [FREE Full text] [doi: 10.1016/j.jsat.2010.01.012] [Medline: 20307801]

22. Carroll KM, Weiss RD. The role of behavioral interventions in buprenorphine maintenance treatment: a review. Am J Psychiatry 2017;174(8):738-747 [FREE Full text] [doi: 10.1176/appi.ajp.2016.16070792] [Medline: 27978771]

23. Brooklyn JR, Sigmon SC. Vermont hub-and-spoke model of care for opioid use disorder: development, implementation, and impact. J Addict Med 2017;11(4):286-292 [FREE Full text] [doi: 10.1097/ADM.0000000000000310] [Medline: 28379862]

24. Kawasaki S, Francis E, Mills S, Buchberger G, Hogentogler R, Kraschnewski J. Multi-model implementation of evidence-based care in the treatment of opioid use disorder in Pennsylvania. J Subst Abuse Treat 2019;106:58-64 [FREE Full text] [doi: 10.1016/j.jsat.2019.08.016] [Medline: 31540612]

25. Bickel WK, Marsch LA, Buchhalter AR, Badger GJ. Computerized behavior therapy for opioid-dependent outpatients: a randomized controlled trial. Exp Clin Psychopharmacol 2008;16(2):132-143 [FREE Full text] [doi: 10.1037/1064-1297.16.2.132] [Medline: 18489017]

26. Christensen DR, Landes RD, Jackson L, Marsch LA, Mancino MJ, Chopra MP, et al. Adding an internet-delivered treatment to an efficacious treatment package for opioid dependence. J Consult Clin Psychol 2014;82(6):964-972 [FREE Full text] [doi: 10.1037/a0037496] [Medline: 25090043]

27. reSET-O/reset: how it works. Pear Therapeutics. URL: https://www.resetforrecovery.com/how-it-works [accessed 2020-06-30]

28. Gustafson DH, McTavish FM, Chih MY, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]

29. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SL, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002;32(6):959-976. [doi: 10.1017/s0033291702006074] [Medline: 12214795]

30. Andrews G, Slade T. Interpreting scores on the Kessler Psychological Distress Scale (K10). Aust N Z J Public Health 2001;25(6):494-497. [doi: 10.1111/j.1467-842x.2001.tb00310.x] [Medline: 11824981]

31. Lang AJ, Stein MB. An abbreviated PTSD checklist for use as a screening instrument in primary care. Behav Res Ther 2005;43(5):585-594. [doi: 10.1016/j.brat.2004.04.005] [Medline: 15865914]

32. Lang AJ, Wilkins K, Roy-Byrne PP, Golinelli D, Chavira D, Sherbourne C, et al. Abbreviated PTSD checklist (PCL) as a guide to clinical response. Gen Hosp Psychiatry 2012;34(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2012.02.003] [Medline: 22460001]

33. Weather FW, Litz BT, Herman DS, Huska JA, Keane TM. The PTSD checklist (PCL): reliability, validity, and diagnostic utility. In: 9th Annual Convention of the International Society for Traumatic Stress Studies. 1993 Presented at: ISTSS '93; November 13-15, 1993; San Antonio, TX.

34. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

35. Sobell LC, Sobell MB. Timeline follow-back: a technique for assessing self-reported alcohol consumption. In: Litten RZ, Allen JP, editors. Measuring alcohol consumption: psychosocial and biological methods. Totowa, NJ: Humana Press; 1992:41-72.

36. Sobell MB, Sobell LC. An excellent springboard. Addiction 2000;95(11):1712-1715. [Medline: 11219379]

37. Petry NM, Peirce JM, Stitzer ML, Blaine J, Roll JM, Cohen A, et al. Effect of prize-based incentives on outcomes in stimulant abusers in outpatient psychosocial treatment programs: a national drug abuse treatment clinical trials network study. Arch Gen Psychiatry 2005;62(10):1148-1156. [doi: 10.1001/archpsyc.62.10.1148] [Medline: 16203960]

38. Budney AJ, Higgins ST. Therapy manuals for drug addiction: a community reinforcement plus vouchers approach: treating cocaine addiction. Rockville, MD: National Institute on Drug Abuse; 1998.

39. Stitzer M, Petry N. Contingency management for treatment of substance abuse. Annu Rev Clin Psychol 2006;2:411-434. [doi: 10.1146/annurev.clinpsy.2.022305.095219] [Medline: 17716077]

Abbreviations

CM: contingency management
CRA: Community Reinforcement Approach
IAFF: Intervention Acceptability Feedback Form
MOUD: medication for opioid use disorder
OTP: opioid treatment program
OUD: opioid use disorder
PTSD: posttraumatic stress disorder
SUD: substance use disorder
TES: Therapeutic Education System
TLFB: Timeline Follow-Back
WIAAF: Weiner Intervention Acceptability, Appropriateness, and Feasibility form

Original Paper

A Physical Activity Just-in-time Adaptive Intervention Designed in Partnership With a Predominantly Black Community: Virtual, Community-Based Participatory Design Approach

Maria Cielito Robles[1], BS; Mark W Newman[2], PhD; Aalap Doshi[3], MSc; Sarah Bailey[4], MS; Linde Huang[3], BS; Soo Ji Choi[2], BFA; Chris Kurien[2], MS; Beza Merid[5], PhD; Joan Cowdery[6], PhD; Jessica R Golbus[5], MS, MD; Christopher Huang[5], BA; Michael P Dorsch[7], MS, PharmD; Brahmajee Nallamothu[5], MPH, MD; Lesli E Skolarus[1], MS, MD

[1]Department of Neurology, University of Michigan, Ann Arbor, MI, United States

[2]School of Information, University of Michigan, Ann Arbor, MI, United States

[3]Michigan Institute for Clinical & Health Research, University of Michigan, Ann Arbor, MI, United States

[4]Bridges into the Future, Flint, MI, United States

[5]Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States

[6]School of Health Promotion and Human Performance, Eastern Michigan University, Ypsilanti, MI, United States

[7]College of Pharmacy, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Lesli E Skolarus, MS, MD
Department of Neurology
University of Michigan
1500 East Medical Center Drive
Ann Arbor, MI, 48109
United States
Phone: 1 7349369075
Email: lerusche@umich.edu

Abstract

Background: Black people are disproportionally impacted by hypertension. New approaches for encouraging healthy lifestyles are needed to reduce hypertension and promote health equity in Black communities.

Objective: In this report, we describe the early-stage, virtual design of a just-in-time adaptive intervention (JITAI) to increase physical activity in partnership with members of a low-income, predominantly Black community.

Methods: The hallmark of JITAIs is highly contextualized mobile app push notifications. Thus, understanding participants' context and determinants of physical activity are critical. During the height of the COVID-19 pandemic, we conducted virtual discovery interviews and analysis guided by the Behavior Change Wheel (which focuses on participants' capacity, opportunity, and motivation to engage in physical activity), as well as empathy mapping. We then formed a community-academic participatory design team that partnered in the design sprint, storyboarding, and paper prototyping.

Results: For this study, 5 community members participated in the discovery interviews, 12 stakeholders participated in the empathy mapping, 3 community members represented the community on the design team, and 10 community members provided storyboard or paper prototyping feedback. Only one community member had used videoconferencing prior to partnering with the academic team, and none had design experience. A set of 5 community-academic partner design principles were created: (1) keep users front and center, (2) tailor to the individual, (3) draw on existing motivation, (4) make physical activity feel approachable, and (5) make data collection transparent yet unobtrusive. To address community-specific barriers, the community-academic design team decided that mobile app push notifications will be tailored to participants' baseline mobility level and community resources (eg, local parks and events). Push notifications will also be tailored based on the day (weekday versus weekend), time of day, and weather. Motivation will be enhanced via adaptive goal setting with supportive feedback and social support via community-generated notifications.

Conclusions: We completed early-stage virtual design of a JITAI in partnership with community participants and a community design team with limited design and videoconferencing experience. We found that designing JITAIs with the community enables these interventions to address community-specific needs, which may lead to a more meaningful impact on users' health.

Introduction

Hypertension is the most important modifiable risk factor for cardiovascular disease, the leading cause of mortality in the United States [1,2]. Black Americans have the highest prevalence of hypertension of any racial/ethnic group in the United States, contributing to their disproportionately high burden of cardiovascular disease [3]. Lifestyle modifications, such as increasing physical activity, play an essential role in hypertension management [4,5]. However, about half of Americans do not meet the American Heart Association recommendation of 150 minutes/week of moderate-intensity physical activity [6], and Black Americans engage in less physical activity than White Americans [6-9]. To improve health equity, new approaches to promote healthy lifestyles are needed.

One approach may be just-in-time adaptive interventions (JITAIs, pronounced as "jedis") [10-12]. This novel approach incorporates real-time data streams from wearable devices (eg, Apple Watch, Fitbit) to generate tailored notifications delivered at moments when there is a higher likelihood of their successful adoption. Not surprisingly, JITAIs introduce several design considerations, including the timing, frequency, modality, and presentation of intervention support, along with an understanding of the impact of behavior change strategies such as "provide feedback on performance" and "prompt specific goal setting" [13]. In the myBPmyLife project, as part of the Wearables In Reducing risk and Enhancing Daily Lifestyle (WIRED-L) center, we are designing and adapting a JITAI to increase physical activity and reduce blood pressure that will then be tested in a randomized controlled trial [11].

The myBPmyLife JITAI is based on a JITAI designed and tested among a predominantly White population [11]. To ensure that JITAI design is inclusive of a low-income, predominantly Black community, we undertook a community-based participatory design approach [14,15]. This paper describes our virtual, early-stage, community participatory approach, as necessitated by the COVID-19 pandemic, grounded in health behavior theory, to adapt a physical activity JITAI to be community-inclusive.

Methods

Community-Based Participatory Design Approach

Flint, Michigan, has a population of 95,358 people: 54% are Black people, and 12% of adults have a bachelor's degree or higher [16]. About 40% of Flint residents self-report hypertension [17]. Furthermore, Flint is recovering from a lead crisis, which may be associated with the development of hypertension [18,19].

Our work in Flint is based on an established community-academic partnership focused on addressing racial disparities in cardiovascular disease, which has been in place for over a decade [20]. The principal community-partner organization is Bridges into the Future, a grassroots community organization dedicated to the health and well-being of the Flint community. The principal academic partner organization is the University of Michigan-Ann Arbor, with support from the University of Michigan-Flint. Before starting the design phase, the myBPmyLife project underwent review and received approval from the Flint Community Ethics Review Board (HUM00181363) [21]. Bridges into the Future led recruitment for discovery interviews and empathy mapping. This involved identifying possible participants with a smartphone or computer access, a willingness to use videoconferencing, and initial interest. Contact information was then provided to the academic team, who obtained consent and delivered videoconferencing training. The community design team was composed of the director of Bridges into the Future (SB) as well as 2 other community members recruited by Bridges into the Future to provide a wider view of the community. The director of Bridges into the Future (SB) and one of the academic partners (LES) led the community advisory board, which consists of representatives from the Flint medical community, including Hamilton Health Center, the largest Federally Qualified Health Center in Flint, and the community-based organization partners. The community partners and community advisory board will continue to partner in designing the app and trial and will lead community dissemination of the results.

Design Process Overview and Formation of Community Design Team

Our early-stage design consisted of five components: (1) discovery interviews, (2) empathy mapping, (3) design sprint, (4) storyboarding, and (5) paper prototyping (Figure 1). Only one of the community participants had used videoconferencing prior to the discovery interviews or empathy mapping. After completion of empathy mapping, the academic and community partners determined that technology was a barrier to creativity and innovation. Thus, we formed a community participatory design team, whose members underwent more intensive training on the use of Zoom, a videoconferencing platform, and intensive training with Miro, a virtual collaborative whiteboard platform. Miro and Zoom were selected based on their ease of use and compatibility with phones. We then moved to design sprints, storyboarding, and paper prototyping. The community design team consisted of three Black women aged 32-72 years who provided near real-time, longitudinal feedback.

Figure 1. Discovery and design. COM-B: Capability, Opportunity, and Motivation.
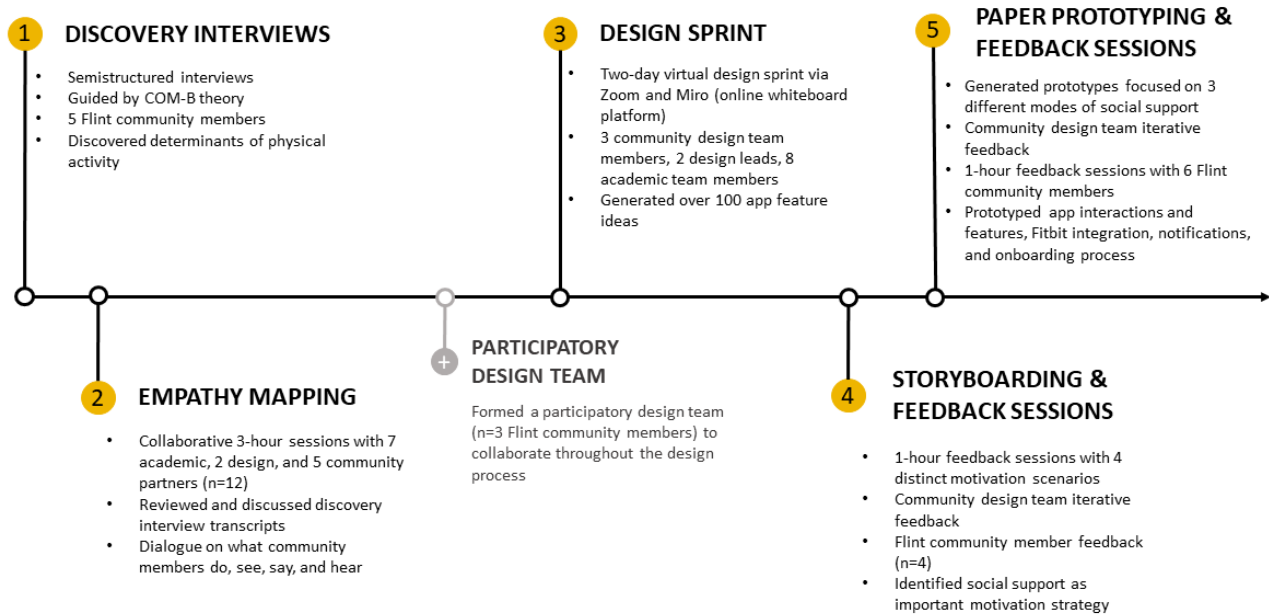
Discovery Interviews

We conducted virtual semistructured interviews with Flint adults with hypertension from July to August 2020. The community partners identified individuals interested in the study. The academic team then consented eligible participants and conducted an individualized training session on using the Health Insurance Portability and Accountability Act–approved videoconferencing service Zoom. Our interview guide was based on the Capability, Opportunity, and Motivation (COM-B) model. The Behavior Change Wheel (BCW), which synthesizes 19 established frameworks of behavior change, posits that for individuals' behaviors to change, they must have the capability, opportunity, and motivation to enact the behavior. The findings from the COM-B–guided interviews were then mapped to the BCW framework to identify candidate behavior change strategies. We also queried participants about smartphone and wearable device usage. All interviews were conducted via Zoom by a primary interviewer and were recorded and transcribed. For analysis, we used a direct content analysis with coding into the COM-B framework [22].

Empathy Mapping

A collaborative 3-hour empathy mapping session was held with academic partners, designers, and community members who participated in the discovery interviews to contextualize the discovery interviews and gain a collective understanding of end user needs. Empathy mapping allows researchers and designers to hear and understand the end user's struggles and frustrations firsthand and in their own words [23]. Before the session, participants were assigned two transcripts to review, each from the discovery interviews. During the videoconferencing session, participants were asked a series of questions and were instructed to reflect on their perceptions and understanding of the reviewed transcripts. The questions included the following:

1. What does the interview participant think and feel?
2. What does the interview participant see?
3. What does the interview participant hear?
4. What does the interview participant do?
5. What does the interview participant say?

These data were added to the qualitative coding results to enable a deeper understanding of the data.

Design Sprint

Design sprints are typically conducted in person to move from empathy generation to the intervention design phase [24]. The design sprint aimed to generate as many diverse and creative ideas as possible for a physical activity intervention in the Flint community. Unlike traditional in-person design sprints, we used videoconferencing with Zoom and Miro. Before the first session, a 60- to 90-minute individual training session was conducted with community participants, often with in-home support from a family member or friend on using Zoom and Miro. Facilitators provided additional real-time tech support during the design sessions. Despite these efforts, real-time workarounds, including email and text, were still needed during the design sprint.

The first day of the design sprint was dedicated to the following: (1) identifying the primary challenges to physical activity and (2) developing problem-solving statements. After a short icebreaker, talks were delivered by a digital technology expert (on mobile technology), a physical activity behavior change expert, a research assistant (on the process of data collection), and a leader from the Flint community. Participants were then asked to create "How Might We" statements to launch brainstorming sessions.

The second day was focused on brainstorming and discussing mobile app features that would address the identified challenges. It began with a Lightning Demos activity [24]. Lightning Demos are a preliminary design exercise by which individuals identify and present 2-3 product and service inspirations prior to concept sketching. For the concept sketching portion of the design sprint, we elected to use three exercises: (1) Notes, (2) Ideas, and (3) Crazy Eights. For exercise 1 (Notes), the primary goal was to

XSL•FO
RenderX

note concepts and problems we previously discussed and established. During exercise 2 (Ideas), participants were given 15 minutes to elaborate on the ideas. Participants were encouraged to creatively doodle, write sample text headlines, and sketch diagrams, cartoons, and stick figures. The messaging provided was "the crazier the idea, the better!" For the final portion of the brainstorming activity, we used the Crazy Eights exercise, asking participants to create 8 distinct ideas in 8 minutes [24]. Participants then synthesized their ideas into one final concept, which they shared with the group. We discussed the concepts as a group and then voted on the ones we thought would form the core intervention. Following the design sprint, the design team categorized all sketches and notes using Evernote, a digital note organizing tool.

Storyboarding and Paper Prototyping

The design team used the information generated during the sprint to create storyboards. These storyboards were presented iteratively to our community design team and then during 4 videoconferencing feedback sessions with Flint community members. Community participant identification, recruitment, and videoconferencing training mirrored that of the discovery interviews. During each session, community design team members were shown the 4 storyboards sequentially. Participants were asked for their reactions—specifically, which parts of the storyboard resonated for them and which elements warranted change or were missing. Design team members were present for all sessions, which were conducted virtually and recorded. After each storyboard was discussed, participants were asked to rank them based on the likelihood of use. The highest-rated storyboards guided the design and development

of the paper prototypes. A critical insight from storyboard feedback was the importance of social interaction, which we then represented in different forms through paper prototypes of the JITAI. Feedback on the paper prototypes was gathered over videoconferencing iteratively from the community design team and then from 6 community participants.

Results

Our findings are categorized into two main domains: (1) user needs discovery and (2) design.

User Needs Discovery

We conducted 5 discovery interviews and 14 stakeholders (7 academic partners, 5 community members, 2 design team members) participated in an empathy mapping session to further contextualize these findings. Guided by the direct content coding using the COM-B framework and enhanced by empathy mapping, we identified unique community physical activity determinants. Physical capability, often due to more chronic physical impairments, was a barrier to physical activity, as was physical opportunity (eg, access to places for older adults to exercise, gym membership costs, and personal safety). Participants reported a desire to exercise more (reflective motivation). On the other hand, automatic motivation (ie, impulses) was a challenge, with community members lamenting how challenging it can be to adhere to a physical activity plan and feeling discouraged by their failure to adhere. These findings were then linked to the BCW to identify behavior change strategies to address the identified barriers and enhance facilitators (Table 1).

Table 1. Results from discovery interviews and empathy mapping mapped to behavior change techniques through the Behavior Change Wheel.

Theme	Exemplar quotes	Selected intervention functions	Behavior change technique
Physical capability	• *It limits me, very much so, because sometimes I have to walk with a cane with these issues. There's been time when I've fallen, and because I haven't had any physical therapy, I've not been able to get out and do things, my body sort of feels tightened and locked up* [Participant 6] • *I can't run. I can still walk. My knees are not good like they used to be* [Participant 1]	Training • Enablement	• Instruction on how to perform a behavior (level of physical activity); graded tasks • Action planning; social support
Psychological capability	• *I think you should at least spend 40 minutes to an hour every day, doing something. Every day. Even if you have small things around the house, or like a 10-minute exercise, just something, some type of physical activity every day, I think, is a must.* [Participant 4]	• Enablement	• Social support
Social opportunity	• *Not a lot of physical activity unless we're cleaning the house or something like that because everyone will work in the yard in our household but not a lot.* [Participant 2] • *My daughter wanted to take me [to the track] but she doesn't go too much herself* [Participant 3]	• Enablement	• Social support
Physical opportunity	• *The neighborhood has parks but I don't see anything for seniors. The only things would be the old forest park where you could walk around the park. Other than the rest of the park is surrounded by kids or teenagers. They got basketball courts and play structures for the toddlers but they don't have tracks for the seniors so it's not senior friendly in the neighborhood I live in.* [Participant 1] • *I would love to be able to have a gym membership. But there's only a certain amount of income—I'm on a fixed income. If I was able to afford it, I certainly would love to be able to…But everything is so expensive, it's difficult.* [Participant 6]	• Environmental restructuring • Enablement	• Prompts/cues; restructure physical environment (or make aware of options) • Goal setting
Reflective motivation	• *I believe that it is very important. For one, as you get older, things kind of lock up. You want to keep those things loose* [Participant 4]	• Persuasion	• Credible source
Automatic motivation	• *It was easy for me then because I was slim but then I got my stroke and my balance and I couldn't do things and I was sitting home all the time and gained 100 pounds ever since and have struggled ever since* [Participant 3] • *I had a good regimen going and I was doing really good. Then I had to go to the nursing home and hospice more. Then my son went to the rehab center. I wore myself down to the point where it was easier to come home, sit in the couch, and watch NBC and just be snacking.* [Participant 2]	• Persuasion • Incentivization • Environmental restructuring • Enablement	• Credible source • Audit and feedback • Prompts/cues • Goal setting, self-talk

Participants primarily used their smartphones for email and text messages, followed by scheduling, receiving news and weather information, taking and storing photos, and using social media (eg, Facebook). None of the participants reported using wearable technology.

Design

The design sprint was facilitated by two members of the design team and attended by 8 academic team members, 2 design team members, and the 3 community design team members. The design exercises yielded approximately 100 app features. Each feature was analyzed as an artifact and tagged into a categorical theme by the design team. Overall, 204 total tags were categorized into 55 themes. The most common tags were motivation (n=22), options (n=13), personalization (n=13), social connection (n=13), and social motivation (n=13). Five design principles emerged from the design sprint and these guided the remainder of the design process (Textbox 1).

XSL•FO
RenderX

Textbox 1. Design principles generated by the community-academic partnership.

1. Keep the user front and center (users first, then researchers, then technologists).

2. Tailor to the individual (customize features to individual users depending on physical and environmental barriers, as well as personal preferences).

3. Draw on existing motivation (connect activities with the user's existing motivation, such as social connections or personal goals).

4. Make physical activity feel approachable (broaden the definition of "exercise" or "physical activity" to simply "moving one's body" in daily life).

5. Make data collection transparent yet unobtrusive for the user (make it clear to the user what data are being collected, and ensure the data are gathered in a minimally intrusive manner).

Based on our user needs discovery and design sprint results, we focused on motivation. Motivational strategies including gamification, social interaction, education, and emotional support were incorporated into storyboards. Overall, the idea of social interaction as part of their physical activity journey most resonated with the community design team and community participants. The concept of social interaction was then broken down into 3 main concepts for paper prototypes. The first prototype was a 1:1 interaction with an accountability buddy. The second prototype focused on developing teams of people who would challenge each other. The final prototype was community-focused and included participant-facing community statistics and community-gathered suggestions. There were mixed opinions on the "accountability buddy" feature, with one participant asking "what if you don't have anybody?" Two participants noted it might be challenging to create a group of people who have similar availabilities. Lastly, participants viewed the community-focused feature positively, although there was concern about the interpretation of statistics. All were in favor of community-gathered suggestions.

Design Implications

Tailored notifications are the foundation of JITAIs. Our first major design decision was to tailor push notifications to participants' respective communities (Table 2). Second, we identified wide variation in physical capability, necessitating tailoring to a participant's functional capacity. For example,

participants with joint pain will be encouraged to engage in shorter, more frequent bursts of physical activity. One behavior change technique to increase automatic motivation is prompts/cues. However, the prompts/cues must be contextualized and, as such, we will tailor push notifications based on day (weekday versus weekend), time of day, and weather. We will enhance reflective and automatic motivation as well as social opportunity and psychological capability by creating a bank of community-generated push notifications to promote social support and app engagement. Community-generated notifications will infuse community voices into the intervention, leading to enhanced authenticity and relatability. Finally, we will acknowledge and support participants when they do not meet their goals through support and adaptive goal setting of steps [25]. Participants will be assigned a weekly task to review their progress over the previous 7 days and select a daily step goal for the upcoming week. The system will automatically suggest a step goal and allow the user to modify the proposed goal within a specified range before accepting it. The notifications will be celebrative when participants reach their goals and support them when they do not reach their goals [26]. During the trial, all participants will be given a wearable. To assess engagement with the push notifications, we will assess the change in step count in the 60 minutes following the push notification. In the intervention, participants will receive daily messages across 4 time periods.

Table 2. Results of community-based participatory design approach to just-in-time adaptive interventions.

Overall findings from community design	Just-in-time adaptive intervention design implications
Physical opportunity is a barrier	Tailor push notifications based on enrollment community
Expanding definition of physical activity (eg, household chores, walking up/down stairs)	Include expanded definition in materials and notifications
Limited physical capability/mobility (eg, pain or multiple comorbidities)	Tailor push notifications to participants' functional capacity
Social support (ie, strong familial and community ties) is a facilitator of physical activity	Community-generated supportive push notifications
High reflective motivation (want to exercise more, but it is hard)	Community-generated supportive push notifications
Automatic motivation is a barrier to physical activity	Contextual tailoring of push notifications based on time, weather, and step count; weekly adaptive step goals; audit and feedback of steps (daily, weekly, monthly)

Discussion

Partnering with community members who had limited experience with videoconferencing, we completed virtual ideation, design sprint, and early prototyping. Virtual

community-based participatory design was facilitated by the creation of a community design team. Our findings highlight that communities have unique needs. Designs that focus on addressing these community-specific barriers and facilitators may increase the efficacy of the JITAIs and ultimately reduce cardiovascular disease disparities.

We initially planned to include the voices of many community stakeholders to understand the breadth of the community experience and capture community innovation. However, the COVID-19 pandemic posed recruitment challenges, as many of our prepandemic recruitment strategies were predicated on in-person interactions with community partners, and there were technological challenges given the lack of videoconferencing experience among nearly all of our participants. After conducting ideation activities over Zoom, we realized that technology was a barrier to creativity. Thus, we created a 2-stage approach to the early design. We initially obtained iterative, near real-time feedback from the community design team, made adaptations, and then shared later prototypes with community users. We found that this 2-stage approach engendered feelings of responsibility among the community design team, increased trust between the community and design team, and optimized the time spent with community participants.

Our community-based participatory design process had several limitations. Our community design team did not include any men, and men were underrepresented among our community participants more generally. Although a community-based participatory research approach to recruitment was successful, other best practices for recruitment, such as face-to-face interaction, were unable to be initiated due to the COVID-19 pandemic [27-29]. Future studies should use strategies that promote the recruitment of men, such as prioritization of men by the study team, including men as part of the recruitment team, and expanding community partners. Of note, a health emergency prohibited our lead male community partner from participating, although women predominantly lead community-based organizations in Flint. Second, we conducted this early phase design during the COVID-19 pandemic. Pandemic stress may have influenced the study findings, although discovery interviews were framed to prompt participants to consider themselves prepandemic. In addition, given the dependence on videoconferencing during the pandemic, community members may be more facile at using this technology now.

We found it was feasible to conduct a virtual, early-stage, community-based participatory design process with participants with limited videoconferencing experience. Initial design prototypes are currently being developed in tandem with a community-generated push notifications bank. Feedback sessions with the community design team and community members will assess our designs in preparation for clinical trial enrollment.

Acknowledgments

Conflicts of Interest

BN is an an Editor for Circulation:Cardiovascular Quality & Outcomes, a Journal of the American Heart Association. They receive a stipend for this work. The American Heart Association is a sponsor of this study.

References

1. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, National Heart, Lung, and Blood Institute Joint National Committee on Prevention, Detection, Evaluation,Treatment of High Blood Pressure, National High Blood Pressure Education Program Coordinating Committee. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. JAMA 2003 May 21;289(19):2560-2572. [doi: 10.1001/jama.289.19.2560] [Medline: 12748199]
2. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association. Circulation 2019 Mar 05;139(10):e56-e528 [FREE Full text] [doi: 10.1161/CIR.0000000000000659] [Medline: 30700139]
3. Carnethon MR, Pu J, Howard G, Albert MA, Anderson CA, Bertoni AG, et al. Cardiovascular Health in African Americans: A Scientific Statement From the American Heart Association. Circulation 2017 Nov 21;136(21):e393-e423. [doi: 10.1161/cir.0000000000000534]
4. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol 2018 May 15;71(19):e127-e248 [FREE Full text] [doi: 10.1016/j.jacc.2017.11.006] [Medline: 29146535]
5. Lazar RM, Howard VJ, Kernan WN, Aparicio HJ, Levine DA, Viera AJ, et al. A Primary Care Agenda for Brain Health: A Scientific Statement From the American Heart Association. Stroke 2021 Jun;52(6):e295-e308. [doi: 10.1161/str.0000000000000367]
6. Disparities Details by Race and Ethnicity for 2018. Office of Disease Prevention and Health Promotion. URL: https://www.healthypeople.gov/2020/data/disparities/detail/Chart/5069/3/2018 [accessed 2021-05-01]

7.   Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]

8.   Hooker SP, Hutto B, Zhu W, Blair SN, Colabianchi N, Vena JE, et al. Accelerometer measured sedentary behavior and physical activity in white and black adults: The REGARDS study. J Sci Med Sport 2016 Apr;19(4):336-341 [FREE Full text] [doi: 10.1016/j.jsams.2015.04.006] [Medline: 25937313]

9.   Judd SE, Gutiérrez OM, Newby P, Howard G, Howard VJ, Locher JL, et al. Dietary Patterns Are Associated With Incident Stroke and Contribute to Excess Risk of Stroke in Black Americans. Stroke 2013 Dec;44(12):3305-3311. [doi: 10.1161/strokeaha.113.002636]

10. Golbus JR, Dempsey W, Jackson EA, Nallamothu BK, Klasnja P. Microrandomized Trial Design for Evaluating Just-in-Time Adaptive Interventions Through Mobile Health Technologies for Cardiovascular Disease. Circ: Cardiovascular Quality and Outcomes 2021 Feb;14(2):e006760. [doi: 10.1161/circoutcomes.120.006760]

11. Klasnja P, Smith S, Seewald N, Lee A, Hall K, Luers B, et al. Efficacy of Contextually Tailored Suggestions for Physical Activity: A Micro-randomized Optimization Trial of HeartSteps. Ann Behav Med 2019 May 03;53(6):573-582 [FREE Full text] [doi: 10.1093/abm/kay067] [Medline: 30192907]

12. Dorsch MP, Cornellier ML, Poggi AD, Bilgen F, Chen P, Wu C, et al. Effects of a Novel Contextual Just-In-Time Mobile App Intervention (LowSalt4Life) on Sodium Intake in Adults With Hypertension: Pilot Randomized Controlled Trial. JMIR mHealth uHealth 2020 Aug 10;8(8):e16696 [FREE Full text] [doi: 10.2196/16696] [Medline: 32663139]

13. Kramer J, Künzler F, Mishra V, Presset B, Kotz D, Smith S, et al. Investigating Intervention Components and Exploring States of Receptivity for a Smartphone App to Promote Physical Activity: Protocol of a Microrandomized Trial. JMIR Res Protoc 2019 Jan 31;8(1):e11540 [FREE Full text] [doi: 10.2196/11540] [Medline: 30702430]

14. Unertl KM, Schaefbauer CL, Campbell TR, Senteio C, Siek KA, Bakken S, et al. Integrating community-based participatory research and informatics approaches to improve the engagement and health of underserved populations. J Am Med Inform Assoc 2016 Jan;23(1):60-73 [FREE Full text] [doi: 10.1093/jamia/ocv094] [Medline: 26228766]

15. Brewer LC, Fortuna KL, Jones C, Walker R, Hayes SN, Patten CA, et al. Back to the Future: Achieving Health Equity Through Health Informatics and Digital Health. JMIR mHealth uHealth 2020 Jan 14;8(1):e14512 [FREE Full text] [doi: 10.2196/14512] [Medline: 31934874]

16. QuickFacts, Flint city, Michigan. U.S. Census Bureau. 2019. URL: https://www.census.gov/quickfacts/flintcitymichigan [accessed 2021-05-01]

17. Community Assessment for Public Health Emergency response (CASPER) after the Flint water crisis. Centers for Disease Control and Prevention. 2016. URL: https://www.michigan.gov/documents/flintwater/CASPER_Report_540077_7.pdf [accessed 2021-05-01]

18. Harlan WR, Landis JR, Schmouder RL, Goldstein NG, Harlan LC. Blood lead and blood pressure. Relationship in the adolescent and adult US population. JAMA 1985 Jan 25;253(4):530-534. [doi: 10.1001/jama.253.4.530] [Medline: 3968785]

19. Miao H, Liu Y, Tsai TC, Schwartz J, Ji JS. Association Between Blood Lead Level and Uncontrolled Hypertension in the US Population (NHANES 1999-2016). J Am Heart Assoc 2020 Jul 07;9(13):e015533 [FREE Full text] [doi: 10.1161/JAHA.119.015533] [Medline: 32573312]

20. Corches CL, McBride AC, Robles MC, Rehman N, Bailey S, Oliver A, et al. Development, Adaptation and Scale-up of a Community-wide, Health Behavior Theory-based Stroke Preparedness Intervention. Am J Health Behav 2020 Nov 01;44(6):744-755. [doi: 10.5993/ajhb.44.6.1]

21. Carrera J, Key K, Bailey S, Hamm J, Cuthbertson C, Lewis E, et al. Community Science as a Pathway for Resilience in Response to a Public Health Crisis in Flint, Michigan. Social Sciences 2019 Mar 13;8(3):94. [doi: 10.3390/socsci8030094]

22. Manning K. Authenticity in Constructivist Inquiry: Methodological Considerations Without Prescription. Qualitative Inquiry 2016 Jun 29;3(1):93-115. [doi: 10.1177/107780049700300105]

23. Design Kit: The Human-Centered Design Toolkit. IDEO. URL: https://www.ideo.com/work/human-centered-design-toolkit/ [accessed 2020-06-20]

24. Knapp J. Sprint: How to Solve Big Problems and Test New Ideas in Just Five Days. New York, NY: Simon & Schuster; Mar 08, 2016.

25. Korinek EV, Phatak SS, Martin CA, Freigoun MT, Rivera DE, Adams MA, et al. Adaptive step goals and rewards: a longitudinal growth model of daily steps for a smartphone-based walking intervention. J Behav Med 2018 Feb;41(1):74-86. [doi: 10.1007/s10865-017-9878-3] [Medline: 28918547]

26. Deci EL, Eghrari H, Patrick BC, Leone DR. Facilitating internalization: the self-determination theory perspective. J Pers 1994 Mar;62(1):119-142. [doi: 10.1111/j.1467-6494.1994.tb00797.x] [Medline: 8169757]

27. Horowitz CR, Brenner BL, Lachapelle S, Amara DA, Arniella G. Effective recruitment of minority populations through community-led strategies. Am J Prev Med 2009 Dec;37(6 Suppl 1):S195-S200 [FREE Full text] [doi: 10.1016/j.amepre.2009.08.006] [Medline: 19896019]

28. Nicholson LM, Schwirian PM, Groner JA. Recruitment and retention strategies in clinical studies with low-income and minority populations: Progress from 2004-2014. Contemp Clin Trials 2015 Nov;45(Pt A):34-40. [doi: 10.1016/j.cct.2015.07.008] [Medline: 26188163]

29.   Ferris EB, Wyka K, Evenson KR, Dorn JM, Thorpe L, Catellier D, et al. Recruitment and Retention Strategies for Community-Based Longitudinal Studies in Diverse Urban Neighborhoods. JMIR Form Res 2021 Mar 24;5(3):e18591 [FREE Full text] [doi: 10.2196/18591] [Medline: 33759799]

Abbreviations

BCW: Behavior Change Wheel
COM-B: Capability, Opportunity, and Motivation
JITAI: just-in-time adaptive intervention
WIRED-L: Wearables In Reducing risk and Enhancing Daily Lifestyle

XSL•FO
RenderX

Original Paper

Analyzability of Photoplethysmographic Smartwatch Data by the Preventicus Heartbeats Algorithm During Everyday Life: Feasibility Study

Steve Merschel[1], Diplom FH; Lars Reinhardt[2], PhD

[1]Preventicus GmbH, Jena, Germany

[2]Institute for Applied Training Science, Leipzig, Germany

Corresponding Author:
Steve Merschel, Diplom FH
Preventicus GmbH
Ernst-Abbe-Straße 15
Jena, 07743
Germany
Phone: 49 3641 5598450
Email: steve.merschel@preventicus.com

Abstract

Background: Continuous heart rate monitoring via mobile health technologies based on photoplethysmography (PPG) has great potential for the early detection of sustained cardiac arrhythmias such as atrial fibrillation. However, PPG measurements are impaired by motion artifacts.

Objective: The aim of this investigation was to evaluate the analyzability of smartwatch-derived PPG data during everyday life and to determine the relationship between the analyzability of the data and the activity level of the participant.

Methods: A total of 41 (19 female and 22 male) adults in good cardiovascular health (aged 19-79 years) continuously wore a smartwatch equipped with a PPG sensor and a 3D accelerometer (Cardio Watch 287, Corsano Health BV) for a period of 24 hours that represented their individual daily routine. For each participant, smartwatch data were analyzed on a 1-minute basis by an algorithm designed for heart rhythm analysis (Preventicus Heartbeats, Preventicus GmbH). As outcomes, the percentage of analyzable data (PAD) and the mean acceleration (ACC) were calculated. To map changes of the ACC and PAD over the course of one day, the 24-hour period was divided into 8 subintervals comprising 3 hours each.

Results: Univariate analysis of variance showed a large effect ($\eta_p^2 > 0.6$; $P<.001$) of time interval (phase) on the ACC and PAD. The PAD ranged between 34% and 100%, with an average of 71.5% for the whole day, which is equivalent to a period of 17.2 hours. Between midnight and 6 AM, the mean values were the highest for the PAD (>94%) and the lowest for the ACC ($<6\times10^{-3}$ m/s^2). Regardless of the time of the day, the correlation between the PAD and ACC was strong ($r=-0.64$). A linear regression analysis for the averaged data resulted in an almost perfect coefficient of determination ($r^2=0.99$).

Conclusions: This study showed a large relationship between the activity level and the analyzability of smartwatch-derived PPG data. Given the high yield of analyzable data during the nighttime, continuous arrhythmia screening seems particularly effective during sleep phases.

XSL•FO

RenderX

Introduction

Background

Stroke is a leading cause of mortality and disability resulting in considerable economic costs of treatment and poststroke care [1]. With 5.5 million deaths in 2016, stroke was recognized as the second leading cause of death globally, after ischemic heart disease [1]. In the European Union, the number of people living with stroke is estimated to increase by 27% between 2017 and 2047, mostly due to population aging and improved survival rates [2]. Nevertheless, the constantly increasing burden of stroke also implies inadequate implementation or effectiveness of primary prevention strategies [1]. There is no doubt that besides targeting behavioral risk factors, effective screening for conditions that raise stroke risk, such as hypertension, diabetes mellitus, and atrial fibrillation (AF), is essential [1]. In that respect, particular attention must be paid to AF since it represents one of the main causes of stroke [3] and the most common sustained cardiac arrhythmia in adults, affecting between 2% and 4% of the general population [4]. Due to its paroxysmal and often asymptomatic (silent AF) clinical presentation, AF often remains undetected and is frequently only diagnosed upon a suffered stroke [5]. According to the guidelines of the European Society of Cardiology (ESC), 24- to 72-hour electrocardiography (ECG) is the standard of care to detect AF [4]. However, this is usually prescribed only when the patient is already experiencing symptoms (eg, palpitations, exertional shortness of breath, and chest pain), which is a strong indicator that the disease has already progressed to an increased severity (persistent, long-standing persistent, or permanent AF). In addition, the diagnostic yield of standard ECG monitoring is limited in the case of paroxysmal AF since prolonged (>7 days) observation periods are required [5]. Additionally, long-term monitoring using an insertable cardiac monitor is superior to traditional follow-up methods in detecting AF [6]. Thus, in patients over 65 years of age, routine self-monitoring of one's pulse is a class I recommendation in the ESC guidelines for raising the suspicion of AF [4].

Previous Work

Mobile health (mHealth) technologies enabling patient-initiated short-term or continuous long-term pulse recordings are thought to have extraordinary potential to improve the care and management of AF and may yield a practical option to determine AF burden [4,7,8]. In principle, the currently available mHealth devices for AF screening can be divided into the following types: smartphones, smart bands or smartwatches, earlobe sensors, and handheld ECG devices [7]. Aside from handheld ECG devices, all of these devices are based on photoplethysmography (PPG), which is an optical measurement technique for the detection of blood volume changes in the microvasculature [9]. PPG sensors are applied on the surface of the skin, where a light source illuminates the tissue that in turn scatters and partially absorbs the emitted light [10]. A photosensitive diode is used to measure variations in scattered light intensity attributed to cardiac synchronous changes in the blood volume with each heartbeat, resulting in the typical pulsatile component of the PPG waveform [9,10].

The evaluation of the diagnostic performance of various mHealth devices for AF detection revealed the highest sensitivity (95% to 98%) and specificity (95% to 99.6%) for smartphones and their associated applications [7]. Accordingly, both the Preventicus Heartbeats algorithm (PHA, Preventicus GmbH), which is a certified medical device (class IIa, CE marked), as well as the competing product FibriCheck (Qompium NV), showed excellent performance for smartphone recordings [11-14]. However, the diagnostic accuracy of smart bands and smartwatches was highly variable between studies and strongly dependent on the test conditions [7]. This is because PPG measurements are vulnerable to artifacts caused by contact pressure, skin tone, user movement, or bright ambient light [7,9,15]. In addition, the signal can also be affected by physiologic variations such as vasoconstriction, coughing, a deep gasp, or a yawn [10]. Hence, the best data quality is to be expected when subjects are sleeping or sitting still [7]. For 1-minute measurements using wrist-worn devices in a controlled laboratory setting (relaxed sitting position with both arms resting on a firm surface), both algorithms (PHA and FibriCheck) demonstrated sensitivity and specificity comparable to smartphone analyses [16,17].

This Study

Against this background, the question arises whether PPG-based smartwatch measurements provide a sufficient data basis for noninvasive continuous AF screening. The PHA performs an automatic signal quality check before rhythm analysis and outputs the percentage of exploitable minutes [11]. Based on this, the present study aimed to evaluate the analyzability of smartwatch-derived PPG data generated by the PHA during everyday life and to determine the relationship between analyzability and activity level in adults who are in good cardiovascular health. Furthermore, this study aims to determine when it is worthwhile to screen for AF, which will help to save data and increase the life of the device battery.

Methods

Participants and Procedures

A total of 41 (19 female and 22 male) adults with an average age of 43.2 (SD 15.8; range 19-79) years volunteered to participate in this investigation. Excluded from the study were shift workers, those who were unable or unwilling to give informed consent, and patients already diagnosed with a cardiovascular disease, especially heart rhythm disorders (eg, AF or ectopic beats). Each participant was asked to continuously wear a smartwatch for a period of at least 24 hours while following their individual daily routine. Thus, the measurements comprised the whole span of everyday life situations ranging from nonphysical activities, like sleeping or office work, to physically demanding activities, such as sports.

Ethics Approval

The study was conducted according to the Declaration of Helsinki. Informed consent was obtained after verbal and written explanation of the experimental design. This investigation provided the technical preparation for our Clinical Trial "Determine AF Burden with PPG Trial," which was approved

by the Northwest and Central Switzerland Ethics Committee (Project-ID: 2020-01983)

Data Acquisition and Processing

The PPG and mean acceleration (ACC) signal during everyday life were measured at the wrist using a smartwatch (Cardio Watch 287, Corsano Health BV). In total, 3 watches of the same type were used for data collection. Besides a 3D accelerometer (sensor range ±16 g), the device was equipped with a single channel PPG sensor module. A total of 2 green light-emitting diodes (LEDs; peak wavelength of 525 nm and maximum current of 30 mA) served as light sources for the sensor (Figure 1). The smartwatch measured PPG and accelerometric data with a sampling rate of 25 Hz.

Figure 1. The smartwatch used for measuring photoplethysmographic (PPG) and accelerometric data. (A) A photograph of the front side of the wearable used in the study (Cardio Watch 287, Corsano Health BV). (B) The back side of the device indicating the exact position of the light-emitting diode (LED), the PPG sensor, and the charging contacts.

Data transfer and storage was performed via a wireless connection (Bluetooth Low Energy, version 5.0) to a paired and preconfigured smartphone (Galaxy A40, Samsung Electronics Co Ltd) with the Android 10 operating system (Google LLC) that was running a data acquisition application provided by MMT. To avoid data loss, the internal memory (8 MB) of the Corsano smartwatch enabled data buffering for a period of about 24 hours. The PPG and ACC signal were automatically saved as JSON files on the smartphone and manually transferred to a desktop PC after the participant completed the measurement period. The raw data were further processed in MATLAB R2016a (MathWorks Inc). First, all data sets were manually inspected for completeness or potential data gaps. Since the smartwatch did not have an algorithm capable of automatically detecting when the watch was not worn, these phases were eliminated afterwards using custom-written programs in MATLAB. These periods were easy to identify as they appeared as Gaussian white noise in the raw data. Data sets containing more than 10% (2.4 hours) of erroneous data were excluded from the analysis. For each participant, a period of 24 hours was analyzed on a one-minute basis using the PHA. Consequently, a maximum of 1440 minutes (data points) was analyzed for each participant. To map changes over the course of a day, the 24-hour period of a day was divided into 8 equal subintervals (P1 to P8) of 3 hours or 180 minutes. The phases should divide the 24-hour day as sensibly as possible, thus resulting in the following phases: P1 (midnight to 3 AM; sleeping phase), P2 (3 AM to 6 AM; sleeping phase), P3 (6 AM to 9 AM; wake-up phase), P4 (9 AM to noon; working phase), P5 (noon to 3 PM; working phase), P6 (3 PM to 6 PM; working phase), P7 (6 PM to 9 PM; coming to rest phase), and P8 (9 PM to midnight; going to sleep phase). We have assumed that most people tend to behave roughly according to these time windows. Furthermore, on the one hand, one needs a division that is sufficiently large to be suitable for statistical tests. On the other hand, this division should not be too fine because otherwise, one obtains error probabilities that are too small with the Bonferroni correction for multiple testing. For this purpose, a compromise had to be found.

Outcomes

The percentage of analyzable data (PAD) described the ratio between the number of measured PPG minutes and the number of exploitable minutes for the PHA. Exploitable minutes are determined by using the signal-to-noise ratio and the ACC data. For example, data sections disturbed by movement can be detected and marked. If the ratio of disturbed data sections to data sections with sufficient quality is more than 10%, then the

entire minute is nonexploitable. A minute must therefore contain no more than 6 seconds of disturbed data for it to be exploitable [11]. The activity level of the participant (represented by the ACC) was estimated using the triaxial acceleration. For each time increment, the magnitude of acceleration change was calculated using the Euclidean norm (ie, the square root of the sum of squares) (Figure 2). Finally, the standard deviation of this was calculated for each period (P1 to P8) and expressed as the ACC .

Figure 2. A visualization of the method for quantifying the activity level (represented by the ACC) of a participant. The length of the vector between 2 successive triaxial acceleration values (green arrow) was calculated using the Euclidean norm. ACC: mean acceleration.
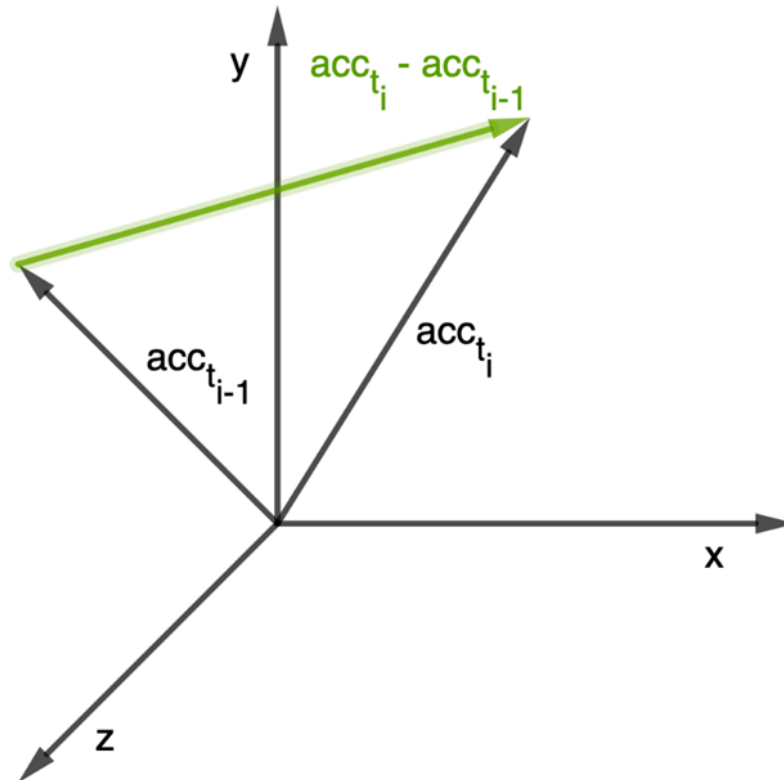
Statistical Analysis

Statistical analyses were performed using SPSS, version 23 (IBM Corp). Results are presented as means with SD and 95% CI. Sex comparisons were verified with Student t tests for unpaired samples. Where appropriate, effect sizes were calculated as standardized mean differences (d) with values >0.2, >0.5, and >0.8, indicating small, moderate, and large effects, respectively. The effects of measurement time on activity level or percent analyzability of PPG data were verified using univariate repeated measurement analysis of variance (ANOVA). Following the findings of the sex comparisons (comparing the outcomes between males and females), the ANOVA was conducted irrespective of sex using the whole sample (N=41). Bonferroni post hoc tests were used to verify localized differences. Practical relevance was estimated using partial eta squared (η_p^2) with values ≥ 0.01, ≥ 0.06, or ≥ 0.14 indicating small, moderate, or large effects, respectively [18]. Pearson product-moment correlations were calculated to examine linear associations between activity level and the analyzability of PPG data. The following criteria were adopted for interpreting the magnitude of correlations (r) between measures: <0.2, trivial; 0.2 to 0.3, small; 0.3 to 0.5, moderate; 0.5 to 0.7, large; 0.7 to 0.9, very large; and 0.9 to 1.0, almost perfect.

Results

The comparisons of means did not reveal differences between sexes (P>.10) and for the PAD (P>.30) in any of the time intervals (P1 to P8). The only effect on a moderate level (d=0.52) was found for ACC in P5. No effects (for ACC d<0.19) were found for the PAD in P1, P2, P5, and P6, as well as for ACC in P4. All remaining sex comparisons resulted in small effect sizes (0.20<d<0.45).

In principle, the temporal changes in the PAD and ACC followed a characteristic but opposite trend throughout the day (Figure 3A and B) that roughly showed a u-shaped (PAD) or inverted u-shaped (ACC) pattern, respectively. The univariate ANOVA showed a large effect (η_p^2>0.6; P<.001) of the time interval on ACC and the PAD (Table 1). Accordingly, ACC values were significantly lower (P<.01) in P1 and P2 than in the remaining intervals (P3 to P8). This resulted in a PAD above 94% between midnight and 6 AM (P1 to P2; Table 1). In the time frame between 9 AM and 9 PM (P4 to P7), ACC values reached a plateau without significant changes (Table 1) on the highest level across the day (Figure 3A). During this period, the ACC average was approximately 37×10^{-3} m/s^2. In accordance with ACC, there was little change in the PAD between 9 AM and 9 PM (P4 to P7). During this phase, values for this percentage ranged between 55% and 60%, on average.

At the same time, these values represented the minimum over the span of a day. Moreover, the values of this parameter in P3 and P8 were clearly higher (73% to 80%), though not reaching the level of P1 and P2.

Interindividual variations in the PAD (SD<7%) and ACC (SD<5×10^{-3} m/s^2) were lowest at night (P1 to P2) and highest around midday (P3 to P4; SD [ACC] 15×10^{-3} m/s^2 to 18×10^{-3} m/s^2; SD [PAD] 10% to 15%).

The PAD and ACC showed an inverse correlation in all investigated periods (Figure 4A). Except for P2 (r=–0.37), all correlations were large (P1, P4, P5, P7) or very large (P3, P6, P8). The linear regression analysis of the averaged data resulted in an almost perfect coefficient of determination (r^2=0.99; Figure 4B).

The values of the PAD ranged between 34% and 100%, with an average of 71.5% for the whole day, which is equivalent to a period of 17.2 hours.

Figure 3. Activity level and analyzability of photoplethysmographic (PPG) data during a 24-hour period representing everyday life with respect to sex. (A) The average standard deviation of the acceleration (ACC). (B) The percentage of PPG data that could be analyzed for arrhythmia screening by the Preventicus Heartbeats algorithm. In both panels, mean values and 95% confidence intervals are symbolized by error bars. Data for males (n=22) and females (n=19) are depicted by filled or blank circles, respectively. The grey dotted lines represent the sex-independent mean.
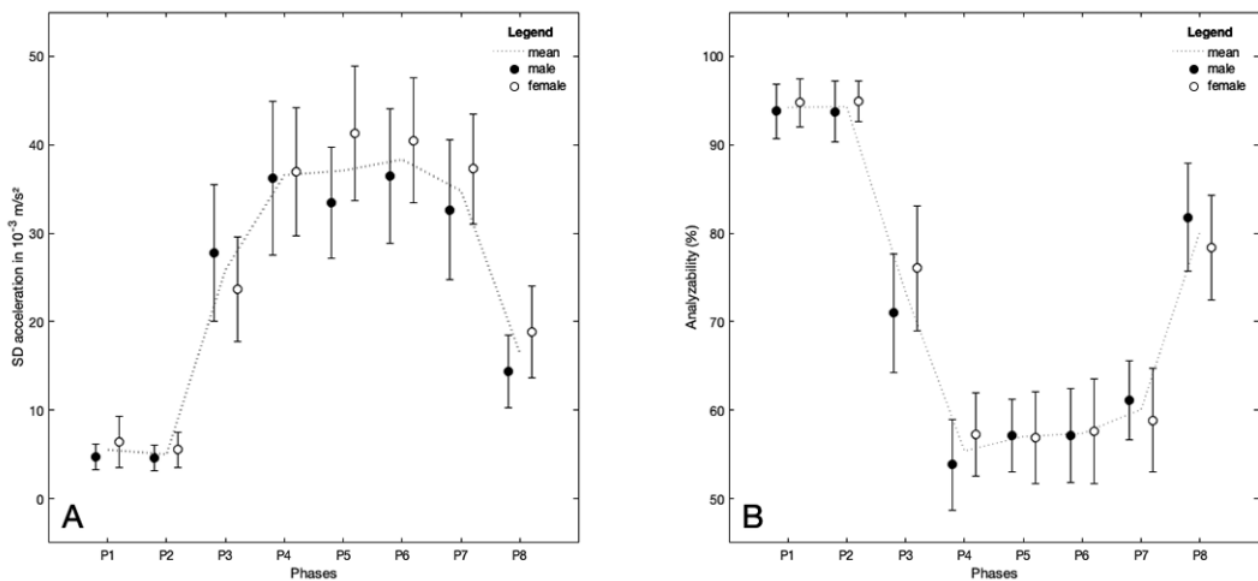
Table 1. The average standard deviation of the acceleration and percentage of analyzable data during each phase of the day (N=41).

Outcome	Phase, mean (SD)								P value[a]	η_p^2	Bonferroni post hoc comparisons (P<.01)[b]
	P1[c]	P2[d]	P3[e]	P4[f]	P5[g]	P6[h]	P7[i]	P8[j]			
ACC[k] in 10^{-3} m/s^2	5.50 (4.70)	5.00 (3.70)	25.9 (15.3)	36.6 (17.4)	37.1 (15.3)	38.3 (16.0)	34.8 (15.8)	16.4 (10.1)	<.001	0.61	P1<P3; P1<P4; P1<P5; P1<P6; P1<P7; P1<P8; P2<P3; P2<P4; P2<P5; P2<P6; P2<P7; P2<P8; P3<P5; P3<P6; P4>P8; P5>P8; P6>P8; P7>P8
PAD[l]	94.2 (6.30)	94.3 (6.50)	73.3 (14.9)	55.4 (10.7)	57.0 (9.82)	57.4 (12.0)	60.1 (11.0)	80.2 (13.2)	<.001	0.75	P1>P3; P1>P4; P1>P5; P1>P6; P1>P7; P1>P8; P2>P3; P2>P4; P2>P5; P2>P6; P2>P7; P2>P8 P3>P4; P3>P5; P3>P6; P3>P7; P4<P8; P5<P8; P6<P8; P7<P8

[a]P value from the univariate analysis of variance.

[b]P<.01 for the comparison of each pair of phases.

[c]P1: midnight to 3 AM.

[d]P2: 3 AM to 6 AM.

[e]P3: 6 AM to 9 AM.

[f]P4: 9 AM to noon.

[g]P5: noon to 3 PM.

[h]P6: 3 PM to 6 PM.

[i]P7: 6 PM to 9 PM.

[j]P8: 9 PM to midnight.

[k]ACC: averaged standard deviation of the acceleration.

[l]PAD: percentage of analyzable data.

Figure 4. The relationship between activity level as expressed by the average standard deviation of the acceleration (ACC) and the analyzability of the photoplethysmographic data for arrhythmia screening generated by the Preventicus Heartbeats algorithm over 8 periods of a single day (P1-P8, compare to Figure 3). (A) Data from different time intervals are symbolized by diamonds, squares, and triangles, and Pearson product-moment correlations (r) are given for each period. (B) Mean values across all subjects for the ACC versus percentage of analyzable data (PAD) in each of the 8 periods. The dashed line represents the linear regression and the corresponding equation with the coefficient of determination (r^2) is given.
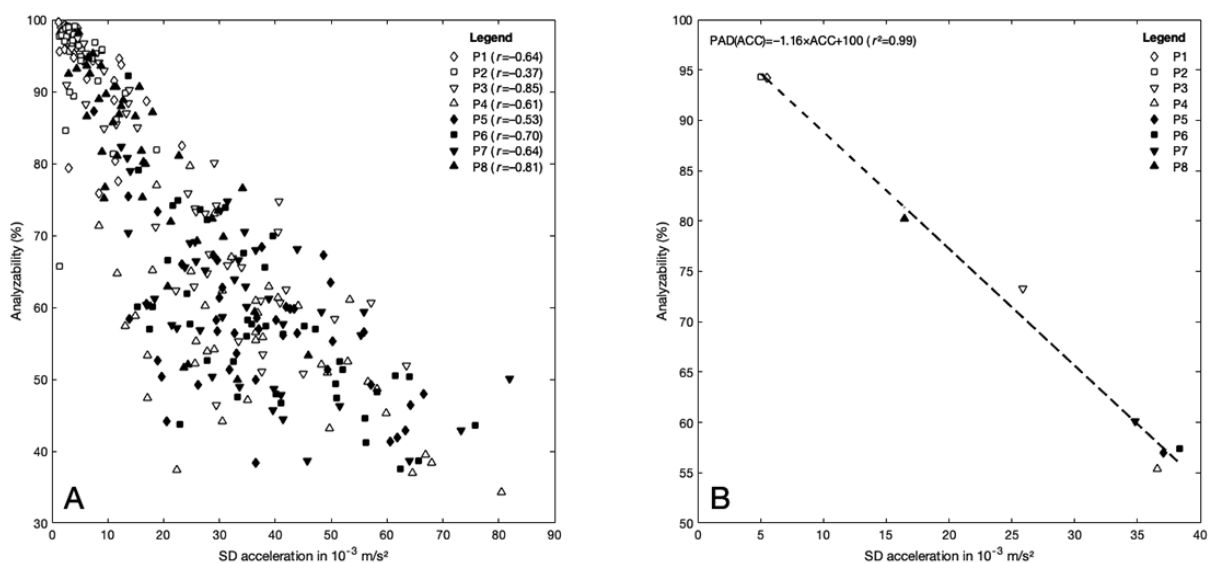
Discussion

Principal Findings

The present study demonstrates a strong relationship between activity level during a routine day and the analyzability of smartwatch-derived PPG data. In addition, a large effect of daytime on the ACC and PAD was found that was mainly based on differences between day and night.

The inverted u-shaped ACC pattern (Figure 3A) represents the rest-activity circadian rhythm of the persons investigated, which reflects the function of the circadian timing system [19]. Since people working in shifts were excluded from this study, it can be reasonably expected that all participants were sleeping and therefore hardly moving between midnight and 6 AM (P1, P2). In accordance with that, the average ACC values and the corresponding variations were minimal. By contrast, the ACC was the highest between 9 AM and 9 PM, which is equivalent to the main activity phase of the day for most people and typically comprises daily working as well as leisure activities. The comparatively high variability in the ACC data during this period can be explained by intraindividual differences in the load level and the temporal course of the everyday activities performed. Furthermore, in the 2 intervals (P3, P8), on an intermediate ACC level, differences in the times for waking up and going to bed are reflected. Since an almost perfect inverse linear relationship (r^2=0.99) has been found between the ACC and the PAD (Figure 4B), all aforementioned statements for ACC also apply in an inverse manner for the PAD.

In line with other studies, it can therefore be concluded that movement strongly impacts the quality of the PPG signal [9,10,15]. During ambulatory monitoring, various types of motion (eg, walking, stretching, and finger tapping) are present that can cause periodic or nonperiodic motion artifacts (MA) [20]. Besides whole-body movements, the relative movement between the PPG sensor and human skin is a potential source of MA, as well. Therefore, it is essential to tighten the wristband of the smartwatch in such a way that the device cannot shift, but keep it loose enough to avoid cutting off blood circulation. Although all participants were instructed to do so in this investigation, it cannot be ruled out that in some cases, the device was not fitted correctly. This may be a potential explanation for the occurrence of some erroneous measurements during the night as well, besides nightly physical movement like going to the bathroom.

The reduction of MA remains a major challenge in processing PPG data since they can be in the same frequency range as the heart rate signal [21]. Although various signal processing methods have been proposed, satisfactory performance in removing or reducing MA has not yet been achieved [21]. Thus, efforts have been made to improve the performance of noise reduction algorithms through adaptations in sensor design.

Nowadays, multichannel PPG measurement systems comprising multiple sensor modules with LEDs of different wavelengths or colors (ie, green, red, or infrared), constitute the standard across multiple models of current PPG-based smartwatches (eg, Polar Vantage series, Samsung Galaxy Watch3, Apple Watch Series 6, and Withings ScanWatch) [21]. Given this, the sensor module used in this study comprised of only 2 LEDs with one wavelength and one photodiode fails to comply with the current standard. Consequently, the PAD is likely significantly higher when measured with newer smartwatches. This hypothesis must be investigated in further studies.

However, the present findings identified a nocturnal time frame of 6 hours with an excellent PAD. Of the 360 minutes between midnight and 6 AM, on average, only about 20 minutes were not exploitable for the PHA, suggesting that AF screening may be particularly effective during this phase. This strategy is supported by several studies showing that AF typically occurs at night or in the early morning hours [22-25].

Limitations

Some limitations must be mentioned. Firstly, the participants were mainly recruited from occupational groups performing office work. We assume that the participants have similar activity profiles (ie, sleep during P1 and P2, wake up during P3, active during P4 to P7, go to sleep during P8). The participants did not keep a diary detailing their activities. Furthermore, no actigraphy techniques were used, either. Other occupational groups such as craftspeople or postal workers might show clearly different activity patterns during the same time of day, which might have an impact on the analyzability of PPG data. In addition, our results only apply for weekdays and not for weekends, when activity profiles most likely look different, especially among the working population. Finally, this study only involved healthy participants and excluded patients with cardiovascular diseases. Hence, the present findings cannot be generalized to the overall older adult population. The analyzability of smartwatch-derived PPG data for a population with major disorders requires further investigation.

Conclusions

This study showed a strong relationship between activity level and the analyzability of smartwatch-derived PPG data and a large effect of time of day on both parameters. That effect was mainly based on the differences between day and night. In conclusion, the present findings suggest that nocturnal AF screening may be particularly effective since the yield of analyzable data was the highest in the time interval between midnight and 6 AM. During this phase, around 94% of the PPG recordings had an appropriate signal quality for rhythm analysis, which is a crucial prerequisite for reliable screening for cardiac arrhythmias such as AF.

Conflicts of Interest

SM is employed full-time as a research developer at Preventicus GmbH, the company involved in the development of the Preventicus Heartbeats algorithm (PHA).

References

1. GBD 2016 Stroke Collaborators. Global, regional, and national burden of stroke, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2019 May;18(5):439-458 [FREE Full text] [doi: 10.1016/S1474-4422(19)30034-1] [Medline: 30871944]

2. Wafa HA, Wolfe CDA, Emmett E, Roth GA, Johnson CO, Wang Y. Burden of stroke in Europe. Stroke 2020 Aug;51(8):2418-2427. [doi: 10.1161/strokeaha.120.029606]

3. Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, et al. Worldwide epidemiology of atrial fibrillation. Circulation 2014 Feb 25;129(8):837-847. [doi: 10.1161/circulationaha.113.005119]

4. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). Revista Española de Cardiología (English Edition) 2021 May;74(5):437. [doi: 10.1016/j.rec.2021.03.009]

5. Dilaveris PE, Kennedy HL. Silent atrial fibrillation: epidemiology, diagnosis, and clinical impact. Clin Cardiol 2017 Jun 08;40(6):413-418 [FREE Full text] [doi: 10.1002/clc.22667] [Medline: 28273368]

6. Sanna T, Diener H, Passman RS, Di Lazzaro V, Bernstein RA, Morillo CA, CRYSTAL AF Investigators. Cryptogenic stroke and underlying atrial fibrillation. N Engl J Med 2014 Jun 26;370(26):2478-2486. [doi: 10.1056/NEJMoa1313600] [Medline: 24963567]

7. Lopez Perales CR, Van Spall HGC, Maeda S, Jimenez A, Laţcu DG, Milman A, et al. Mobile health applications for the detection of atrial fibrillation: a systematic review. Europace 2021 Jan 27;23(1):11-28 [FREE Full text] [doi: 10.1093/europace/euaa139] [Medline: 33043358]

8. Turakhia MP, Kaiser DW. Transforming the care of atrial fibrillation with mobile health. J Interv Card Electrophysiol 2016 Oct;47(1):45-50 [FREE Full text] [doi: 10.1007/s10840-016-0136-3] [Medline: 27306552]

9. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 2007 Mar;28(3):R1-39. [doi: 10.1088/0967-3334/28/3/R01] [Medline: 17322588]

10. Nilsson L. Respiration signals from photoplethysmography. Anesth Analg 2013 Oct;117(4):859-865. [doi: 10.1213/ANE.0b013e31828098b2] [Medline: 23449854]

11. Brasier N, Raichle CJ, Dörr M, Becke A, Nohturfft V, Weber S, et al. Detection of atrial fibrillation with a smartphone camera: first prospective, international, two-centre, clinical validation study (DETECT AF PRO). Europace 2019 Jan 01;21(1):41-47 [FREE Full text] [doi: 10.1093/europace/euy176] [Medline: 30085018]

12. Koenig N, Seeck A, Eckstein J, Mainka A, Huebner T, Voss A, et al. Validation of a new heart rate measurement algorithm for fingertip recording of video signals with smartphones. Telemed J E Health 2016 Aug;22(8):631-636. [doi: 10.1089/tmj.2015.0212] [Medline: 26938673]

13. Krivoshei L, Weber S, Burkard T, Maseli A, Brasier N, Kühne M, et al. Smart detection of atrial fibrillation. Europace 2017 May 01;19(5):753-757 [FREE Full text] [doi: 10.1093/europace/euw125] [Medline: 27371660]

14. Proesmans T, Mortelmans C, Van Haelst R, Verbrugge F, Vandervoort P, Vaes B. Mobile phone-based use of the photoplethysmography technique to detect atrial fibrillation in primary care: diagnostic accuracy study of the FibriCheck app. JMIR Mhealth Uhealth 2019 Mar 27;7(3):e12284 [FREE Full text] [doi: 10.2196/12284] [Medline: 30916656]

15. Bashar SK, Han D, Hajeb-Mohammadalipour S, Ding E, Whitcomb C, McManus DD, et al. Atrial fibrillation detection from wrist photoplethysmography signals using smartwatches. Sci Rep 2019 Oct 21;9(1):15054 [FREE Full text] [doi: 10.1038/s41598-019-49092-2] [Medline: 31636284]

16. Dörr M, Nohturfft V, Brasier N, Bosshard E, Djurdjevic A, Gross S, et al. The WATCH AF trial: smartWATCHes for detection of atrial fibrillation. JACC Clin Electrophysiol 2019 Feb;5(2):199-208 [FREE Full text] [doi: 10.1016/j.jacep.2018.10.006] [Medline: 30784691]

17. Selder J, Proesmans T, Breukel L, Dur O, Gielen W, van Rossum A, et al. Assessment of a standalone photoplethysmography (PPG) algorithm for detection of atrial fibrillation on wristband-derived data. Comput Methods Programs Biomed 2020 Dec;197:105753 [FREE Full text] [doi: 10.1016/j.cmpb.2020.105753] [Medline: 32998102]

18. Cohen J. Statistical Power Analysis for the Behavioral Sciences, Second Edition. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

19. Innominato PF, Focan C, Gorlia T, Moreau T, Garufi C, Waterhouse J, et al. Circadian rhythm in rest and activity: a biological correlate of quality of life and a predictor of survival in patients with metastatic colorectal cancer. Cancer Res 2009 May 26;69(11):4700-4707. [doi: 10.1158/0008-5472.can-08-4747]

20. Zhang Y, Song S, Vullings R, Biswas D, Simões-Capela N, van Helleputte N, et al. Motion artifact reduction for wrist-worn photoplethysmograph sensors based on different wavelengths. Sensors (Basel) 2019 Feb 07;19(3):673 [FREE Full text] [doi: 10.3390/s19030673] [Medline: 30736395]

21. Lee J, Kim M, Park H, Kim IY. Motion artifact reduction in wearable photoplethysmography based on multi-channel sensors with multiple wavelengths. Sensors (Basel) 2020 Mar 09;20(5):1493 [FREE Full text] [doi: 10.3390/s20051493] [Medline: 32182772]

22. Black N, D'Souza A, Wang Y, Piggins H, Dobrzynski H, Morris G, et al. Circadian rhythm of cardiac electrophysiology, arrhythmogenesis, and the underlying mechanisms. Heart Rhythm 2019 Feb;16(2):298-307 [FREE Full text] [doi: 10.1016/j.hrthm.2018.08.026] [Medline: 30170229]

23.   Andresen D, Brüggemann T. Heart rate variability preceding onset of atrial fibrillation. J Cardiovasc Electrophysiol 1998 Aug;9(8 Suppl):S26-S29. [Medline: 9727672]

24.   Shusterman V, Warman E, London B, Schwartzman D. Nocturnal peak in atrial tachyarrhythmia occurrence as a function of arrhythmia burden. J Cardiovasc Electrophysiol 2012 Jun;23(6):604-611 [FREE Full text] [doi: 10.1111/j.1540-8167.2011.02263.x] [Medline: 22429736]

25.   Mitchell ARJ, Spurrell PAR, Sulke N. Circadian variation of arrhythmia onset patterns in patients with persistent atrial fibrillation. Am Heart J 2003 Nov;146(5):902-907. [doi: 10.1016/S0002-8703(03)00405-8] [Medline: 14597942]

Abbreviations

ACC: mean acceleration
AF: atrial fibrillation
ANOVA: analysis of variance
ECG: electrocardiography
ESC: European Society of Cardiology
LED: light-emitting diode
MA: motion artifacts
mHealth: mobile health
MMT: Manufacture Modules Technologies SA
PAD: percentage of analyzable data
PHA: Preventicus Heartbeats algorithm
PPG: photoplethysmography

XSL•FO
RenderX

Original Paper

A Smartphone-Based Intervention as an Adjunct to Standard-of-Care Treatment for Schizophrenia: Randomized Controlled Trial

S Nassir Ghaemi[1], MPH, MD; Oleksandr Sverdlov[2], PhD; Joris van Dam[1], PhD; Timothy Campellone[3], PhD; Robert Gerwien[3], PhD

[1]Novaris Institutes for Biomedical Research, Cambridge, MA, United States

[2]Novartis Pharmaceuticals Corporation, East Hanover, NJ, United States

[3]Pear Therapeutics, Boston, MA, United States

Corresponding Author:
S Nassir Ghaemi, MPH, MD
Novaris Institutes for Biomedical Research
220 Massachusetts Avenue
Cambridge, MA, 02139
United States
Phone: 1 857 253 8950
Email: nassir.ghaemi@tufts.edu

Abstract

Background: Antipsychotic medications have limited benefits in schizophrenia, and cognitive behavioral therapy may be beneficial as an adjunct. There may be potential for implementing mobile cognitive behavioral therapy–based treatment for schizophrenia in addition to standard antipsychotic medications.

Objective: This study aims to determine whether PEAR-004, a smartphone-based investigational digital therapeutic, improves the symptoms of an acute psychotic exacerbation of schizophrenia when it is added to standard treatments.

Methods: This was a 12-week, multicenter, randomized, sham-controlled, rater-blinded, parallel group proof-of-concept study of 112 participants with moderate acute psychotic exacerbation in schizophrenia. This study was conducted in 6 clinical trial research sites in the United States from December 2018 to September 2019. The primary outcome, change in Positive and Negative Syndrome Scale (PANSS) from baseline to week 12 or the last available visit, was analyzed using the mixed-effects regression model for repeated measures, applied to an intent-to-treat sample.

Results: The total PANSS scores slightly decreased from baseline over the study period in both groups; the treatment difference at day 85 between PEAR-004 and sham was 2.7 points, in favor of the sham (2-sided P=.09). The secondary scales found no benefit, except for transient improvement in depressive symptoms with PEAR-004. Application engagement was good, and patient and clinical investigator satisfaction was high. No safety concerns were observed. There was some evidence of study site heterogeneity for the onboarding processes and directions on PEAR-004 product use at baseline and throughout the study. However, these differences did not affect the efficacy results.

Conclusions: In the largest-to-date randomized, sham-controlled study of a digital therapeutic in schizophrenia, PEAR-004 did not demonstrate an effect on the primary outcome—total PANSS scores—when compared with a nonspecific digital sham control. The secondary and exploratory results also did not demonstrate any notable benefits, except for possible temporary improvement in depressive symptoms. This study provided many useful scientific and operational insights that can be used in the further clinical development of PEAR-004 and other investigational digital therapeutics.

Trial Registration: ClinicalTrials.gov NCT03751280; https://clinicaltrials.gov/ct2/show/NCT03751280

XSL•FO
RenderX

Introduction

Schizophrenia is a common condition [1] treated with standard antipsychotic medications, which can help acute exacerbations of delusions or hallucinations but do not improve the long-term course of the illness. Cognitive behavioral therapy (CBT) has been shown to improve the symptoms of schizophrenia both for delusions and hallucinations (*positive* symptoms) and for apathy and flat affect (*negative* symptoms) and functional status (engagement with employment or social interactions) [2]. However, access to CBT can be difficult based on the cost and availability of psychotherapists [3].

Digital therapeutics (DTx) represents a novel treatment modality in which digital technology systems are used as evidence-based therapeutic interventions to prevent, manage, or treat a medical disorder or disease [4]. DTx products based on CBT can be delivered through smartphone apps, and they can potentially provide safe, inexpensive, easy-to-use, consistent quality, personalized treatment strategies for patients with medical needs. The potential value of DTx is magnified during the current COVID-19 pandemic, when face-to-face physician visits are problematic and self-management of long-term conditions is becoming more common. There is an increasing interest in developing DTx for mental health disorders [5,6]. However, the evidence base for the clinical effectiveness of such products is still sparse [7,8]. There are natural challenges in designing clinical trials of DTx interventions, including the choice of a control group, blinding, and potentially lower-than-expected patient engagement [9].

Several digital interventions in psychosis have recently been evaluated in randomized controlled trials (RCTs) [10-12]. The experimental treatments in these studies varied in terms of mode of delivery, features or functionalities, and theoretical framework. In addition, there were different control conditions and different degrees of involvement of health professionals. For instance, Actissist (University of Manchester) [10] was developed as a stand-alone self-management mobile app targeting five domains of early psychosis, such as auditory verbal hallucinations, paranoia, perceived criticism, socialization, and cannabis use. It was tested in a small (n=36) proof-of-concept RCT against an active control condition (ClinTouch), which was a self-reporting symptom severity mobile app. Although the study [10] showed evidence of feasibility, acceptability, safety, and indications of beneficial effects after 12 weeks of treatment with Actissist, its findings are limited owing to the small sample size and the fact that engagement with both experimental and control apps was incentivized.

Another mobile app intervention, Personalized Real-Time Intervention for Motivation Enhancement (PRIME) [11], was designed to target the motivational system of young people with recent-onset schizophrenia spectrum disorders by using social reinforcement to engage and sustain goal-directed behavior. The PRIME intervention provided a supportive web-based environment for social interaction, motivational coaching, personalized goal setting in the domains of health or wellness, social relationships, creativity, productivity, and a system to track progress in achieving personal goals. After 12 weeks of PRIME treatment, there was evidence of improvement in the experimental group of several important components of motivational behavior, depression symptoms, defeatist beliefs, and self-efficacy, compared with the waitlist control condition. Despite a relatively small sample size (a total of 43 participants) and the use of a waitlist control that did not allow for assessment of a relative effect of PRIME compared with other mobile treatment approaches, the study [11] provided important evidence of feasibility, acceptability, and potential clinical benefit of a mobile intervention in this patient population. Notably, this study was implemented fully remotely, across the United States, Canada, and Australia.

The Audio Visual Assisted Therapy Aid for Refractory auditory hallucinations (AVATAR) therapy [12] is a computer-assisted intervention designed on the principles of CBT for psychosis, with the specific aim of controlling persistent, distressing auditory verbal hallucinations. The AVATAR therapy is delivered by experienced clinicians and involves the creation of a computerized representation of the entity (Avatar), which is believed to be the source of the voice heard by the patient, and subsequent therapy sessions at which the clinician facilitates a direct dialogue between the patient and the Avatar, with the goal of having the Avatar less hostile and conceding power over the course of therapy. In the RCT [12], which was formally powered (150 patients randomized equally between the AVATAR therapy and the supportive counseling control condition), the AVATAR therapy led to significantly greater reductions in auditory hallucinations, as assessed by the Psychotic Symptoms Rating Scales Auditory Hallucinations total score, compared with the control condition after 12 weeks of treatment; however, there was no between-group difference at 24 weeks follow-up. The study [12] was conducted at a single center and involved experienced therapists, which limits the generalizability of the results to other centers or to delivery by a wider mental health workforce.

FOCUS is another software-based intervention (delivered via mobile devices) designed with input from both treatment providers and patients to optimize both usability and engagement and developed to be used in conjunction with ongoing outpatient treatment [13]. The feasibility, acceptability, and initial efficacy of FOCUS for improving symptoms and treatment engagement in patients with schizophrenia or schizoaffective disorder was established in a 1-month open-label trial [14]. In another study [15], engagement with FOCUS among patients with schizophrenia was measured during a 6-month period following psychiatric hospitalization discharge. Similar to findings from the 1-month feasibility trial, patients with schizophrenia were highly engaged over the course of 6 months (active use on 82% of the weeks during which they had access to the intervention). Taken together, these preliminary findings regarding FOCUS show the promise of prescription DTx for improving symptoms and treatment outcomes in patients with schizophrenia.

Overall, theory-based digital interventions hold promise in psychosis and schizophrenia spectrum disorders. Given the increasing availability and use of smartphones, CBT-based mobile interventions may potentially augment existing standard-of-care pharmacological treatments or target specific

domains of illness by promoting cognitive and behavioral change strategies.

PEAR-004 is being developed as a prescription digital therapeutic delivered via smartphone for schizophrenia patients who are under the care of a qualified health care professional and are on antipsychotic pharmacotherapy. It is intended to deliver multimodal evidence-based neurobehavioral mechanisms of action, which include cognitive restructuring, illness self-management training, and social skills training. If efficacious, it would demonstrate that 24×7 access to evidence-based coping skills, which when added to medications, may improve symptom management and functional outcomes.

This paper reports the results of a randomized, sham-controlled study of PEAR-004 in patients with schizophrenia.

Methods

Study Design

The study was a multicenter, randomized, sham-controlled, rater-blinded, parallel group proof-of-concept trial of participants with schizophrenia (trial registration: ClinicalTrials.gov NCT03751280). Eligible participants were equally randomized on day 1 to receive either PEAR-004 (investigational digital therapeutic) or sham (control) for a period of 12 weeks. Participants in both groups continued to receive clinician-directed standard-of-care for schizophrenia, including pharmacotherapy. The participants returned to the clinic for outpatient visits at week 4 (day 29), week 8 (day 57), and week 12 (day 85). At each visit, standard assessments, including efficacy and safety, were performed according to the assessment schedule. A final follow-up visit was performed at week 16 (day 115).

Study Objectives

The primary objective was to compare the effect of PEAR-004 versus sham, as assessed by the change from baseline to day 85 in the total Positive and Negative Syndrome Scale (PANSS) score. It was hypothesized that the PEAR-004 group would exhibit a greater reduction in the total PANSS score than the sham group. In addition, retention to assigned study treatment (dropout rate), patient engagement data, secondary efficacy outcomes, clinician outcome assessments, and safety and tolerability data were analyzed.

Study Sites

Potentially eligible participants were enrolled in 6 investigational study sites in the United States. The site details are as follows:

- 1001: Collaborative Neuroscience Network, Garden Grove, California. Principal investigator (PI): Dr David Walling.
- 1002: Collaborative Neuroscience Network, Torrance, California. PI: Dr Lara Shirikjian.
- 1003: Pacific Research Partners, Oakland, California. PI: Dr Corinna Gamez.
- 1005: Meridien Research, Maitland, Florida. PI: Dr Andrea Marraffino.
- 1006: Cherry Health, Grand Rapids, Michigan. PI: Dr Eric Achtyes.
- 1008: Albuquerque Neuroscience, Albuquerque, New Mexico. PI: Dr Glenn Dempsey.

Study Participants

The investigators ensured that all participants being considered for the study met the eligibility criteria at screening. The inclusion and exclusion criteria are described in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Signed informed consent obtained before participation in the study
- Healthy men and women aged 18 to 65 years, inclusive, and in good health as determined by medical history, physical examination, and vital signs at screening
- Structured Clinical Interview for DSM-5–based Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5) diagnosis of schizophrenia and a total Positive and Negative Syndrome Scale score ≥60.
- Proficient in English at the 5th grade reading level or higher, in the judgment of the investigator
- Capable of using a mobile device (compatible with PEAR-004) and using common apps, in the judgment of the investigator

Exclusion criteria

- Major change in primary antipsychotic medication in the prior 4 weeks before screening
- Planning to move out of the geographic area within 3 months
- Unable to use English to participate in the consent process, the interventions, or assessments
- Inability to comply with study procedures owing to severe medical conditions or otherwise
- Meet the DSM-5 diagnosis for a current episode of major depression, mania, or hypomania in the past month
- Meet the DSM-5 diagnosis for a current moderate or severe alcohol or cannabis use disorder in the past 2 months
- Meet the DSM-5 diagnosis for a current substance use disorder (other than alcohol or cannabis) in the past 2 months
- Considered high risk for suicidal behavior based on InterSePT Scale for Suicidal Thinking–Plus score at screening, or in the judgment of the investigator
- Previously participated in a clinical study involving PEAR-004

Randomization

Randomization was implemented by means of interactive response technology (IRT) using permuted blocks of size 4, with a targeted allocation ratio of 1:1. An investigator or delegate at a given study site logged on to the IRT system after confirming that the participant fulfilled all the inclusion or exclusion criteria. The IRT assigned the participant to a treatment arm, which was used by the site staff to request a *prescription access code* in a separate study portal. Trained site staff assisted the participants with all treatment onboarding

activities. The participants were required to use their own personal mobile phones for the study. If a participant did not have their own phone, one was provided to them for use during the study. Additional information on the prescription access code and distribution of study treatment is provided in the site operational manuals provided to the sites. If a participant failed to be treated for any reason, the IRT was updated so that the participant was not treated. Treatment arm assignments were recorded in the case report form. Figure 1 shows the CONSORT (Consolidated Standards of Reporting Trials) participant flow through the trial.

Figure 1. Study participant flowchart—CONSORT (Consolidated Standards of Reporting Trials) diagram. Safety analysis set: all subjects who received the study treatment. Intention-to-treat set: all subjects to whom study treatment has been assigned by randomization and who have a baseline observation and at least 1 postrandomization observation for the analysis end point. I/E: inclusion and exclusion.
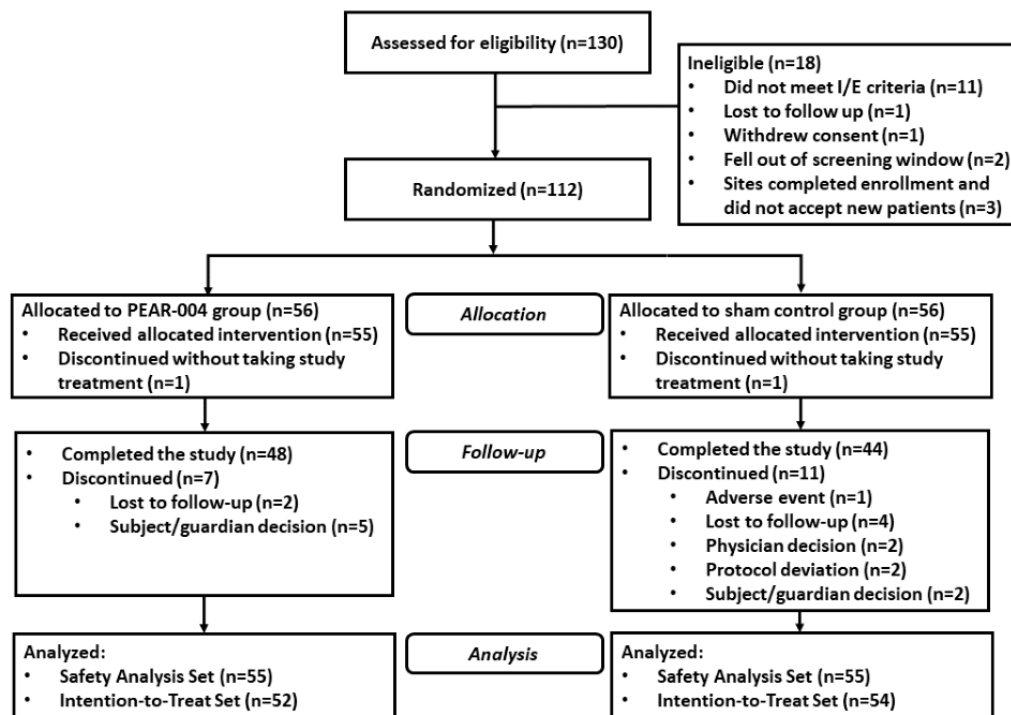
Interventions

On day 1, eligible participants gained access to either PEAR-004 or sham according to their randomization assignment. A single app containing both PEAR-004 and sham was downloaded from the iOS or Android app store to the participant's mobile device, and then the assigned app was unlocked using a prescription access code provided by Pear Therapeutics. The study site staff received training on how to download PEAR-004 or sham to the assigned participant's phone as part of site initiation activities. A single version of PEAR-004 or sham was used for the duration of the clinical study for all randomized participants.

The PEAR-004 smartphone app (iOS and Android-based) was designed as an illness self-management tool. Participant use of PEAR-004 during the treatment period could be either prompted or on demand. Prompted use refers to engagement initiated via 1 of the 3 daily notifications delivered at the following fixed times: 11 AM, 4 PM, and 9 PM. A participant responding to the notification would be brought to a unique survey asking whether the app can be helpful right now. Participants were free to choose from a list of modules or indicate that they were doing

well and did not need help. If they chose a module, they would be brought directly to that module. If they indicated that they were doing well, they would be brought to the PEAR-004 home screen. In addition, PEAR-004 was available on demand. For a particular skill, there would be an option of watching, reading, or listening to content about what the skill is and how they can give it a try. Patients would then view tips for successful practice, and after they practiced a skill, they would be asked to provide feedback on whether the skill was helpful. Helpful skills were stored in the toolbox to promote repeated practice and skill mastery. The final PEAR-004 digital app used in this study was developed from prior versions to have 10 categories of skills: exercise, medication, mindfulness, mood, productivity, sleep, social activity, stress, thoughts, and voices. These categories were informed directly by user research (surveys and interviews) with people with schizophrenia.

In the sham control group, the sham app was downloaded to the participant's phone but did not deliver the active therapeutic content of PEAR-004. Similar to PEAR-004, the sham app delivered 3 daily notifications prompting the participant to open the sham app, and then displayed a prescription timer (*digital*

clock) for the remaining duration of app availability. The sham control arm was chosen to account for the nonspecific effects of engagement with a smartphone. The sham app did not deliver any active coping skills (ingredients of a psychosocial intervention). It appeared similar to the PEAR-004 app in its initial screen and randomization to sham or PEAR-004 was conducted at the time of app download unbeknownst to participants to maintain blinding.

Overall, the similarity and difference between the PEAR-004 app and the sham control app can be summarized as follows:

- Both PEAR-004 and sham sent 3 notifications per day at fixed times—11 AM, 4 PM, and 9 PM—prompting the user to open the app. Once the app was opened, the experience and content differed between the PEAR-004 and sham groups.
- PEAR-004 provided therapeutic content (cognitive and behavioral exercises), whereas the sham app did not.
- Sham had a display of a *digital clock*, whereas the PEAR-004 app did not.
- Neither PEAR-004 nor sham maintained a record of the user's prescription medications or when they were to be taken.

Assessments

Demographics and Baseline Characteristics

Participant demographic and baseline characteristic data were collected from all the study participants. Relevant medical history and current medical conditions present before signing informed consent were recorded.

Engagement With App

Data on patient engagement with the assigned app were collected throughout the study. The cross-platform engagement metrics (for individuals in both PEAR-004 and sham groups) included time using the app, number of days when the app was active, total number of sessions, and number of sessions per day. For participants in the PEAR-004 group, additional metrics were derived, such as the number of skills practiced, number of skills repeated (practiced at least two times), number of skills mastered (practiced at least three times), and number of skills practiced in each of the 10 categories of the PEAR-004 app (exercise, medication, mindfulness, mood, productive, sleep, social, stress, thoughts, and voices).

Efficacy

The following assessments were performed at each study visit:

- PANSS [16]: a 30-item clinician-administered, semistructured interview of schizophrenia symptoms. The PANSS assesses positive (hallucinations, delusions, and thought disorder) and negative (blunted affect, abstract thinking, and general symptomatology). The positive and negative subscale each consist of 7 items rated from 1 (absent) to 7 (extreme) with a minimum score of 7 and maximum score of 49. The general subscale consists of 16 items with a minimum score of 16 and a maximum score of 112. The total PANSS score (positive+negative+general

scores) has a minimum of 30 and a maximum of 210. Higher scores represent greater symptom severity.
- Motivation and Pleasure-Self Report (MAP-SR) [17]: a 15-item self-report that provides a total score index of current motivation or pleasure negative symptoms. MAP-SR includes questions about social pleasure, recreational or work pleasure, close relationships, and motivation and effort to engage in activities of 15 questions with a score of 0-4, summed for a total range of 0-60. Higher values represent better outcomes.
- The Beck Depression Inventory, Second Edition (BDI-II) [18]: a 21-item self-report that provides a total score index of current depression symptom severity. Each item of the BDI-II is scored from 0 to 3, for a total of 0-63. Higher values represent worse outcomes.
- The World Health Organization Quality of Life scale [19]: a 26-item clinician-administered structured interview that assesses psychological functioning and quality of life in four primary domains: social relationships, psychological, physical, and environment. Each of the 26 questions is scored from 1 to 5, and for each of the 4 domains, a total raw score and 2 transformed scores (with ranges of 4-20 and 0-100) are derived. Higher values represent better outcomes.
- Brief Medication Questionnaire [20]: a self-report of medication use, including what medications the participant was currently taking, how they took each medication in the past week, drug effects and bothersome features, and difficulties remembering to take their medication.

Safety

Safety assessments consisted of collecting all adverse events (AEs), serious AEs (SAEs), vital signs, and the InterSePT Scale for Suicidal Thinking–Plus [21,22]. The InterSePT Scale for Suicidal Thinking–Plus is a semistructured interview that assesses the severity of suicidal ideation and behavior, consisting of three parts. Part 1 collects information on 7 days before the visit; there are 13 items that scored 0 (minimum) to 2 (maximum) for suicidality, with a higher score representing a worse outcome. Part 2 collects information on suicidal behavior from the last visit, with nominal categories yes, no, or unknown. Part 3 provides a global rating of status at the time of interview; it is scored 0 (minimum) to 5 (maximum) for suicidality, with a higher score representing a worse outcome. In our study, we focused primarily on the part 3 score (severity of suicidal risk), which was summarized by treatment group and time visit.

Other Assessments

Clinician-reported outcomes included the Clinical Global Impression (CGI) scale [23] and the Clinician Satisfaction Survey assessing the clinician's experience with PEAR-004 and the associated web portal. The CGI consists of the CGI-Severity scale, scored 1 to 7, with larger values indicating greater severity of illness, and the CGI-Improvement scale, scored 1 to 7, with smaller values representing a greater degree of improvement (1=very much improved) and larger values representing a greater degree of worsening (7=very much worse).

Patient-reported outcomes included the Insomnia Severity Index (ISI) [24] and a Subject Satisfaction Survey assessing the participant's experience with PEAR-004 or sham. The ISI includes 7 questions, each scored from 0 to 4, for a total of 0 to 28, where higher values represent more severe insomnia.

Statistical Methods

Sample Size and Power

The required sample size was calculated to address the primary objective of treatment comparison at week 12 with respect to the change in the total PANSS score. Data from 102 participants randomized in a 1:1 ratio to PEAR-004 or sham control would provide 80% power to detect a statistically significant difference between the 2 groups at a 1-sided significance level of 5% assuming the true standardized effect size of 0.5, which is considered a moderate effect size. To account for potential dropouts, 112 participants were enrolled and randomized into the study.

Statistical Analyses

Summary statistics were tabulated for demographics, baseline characteristics, relevant medical histories, and current medical conditions at baseline. The measures of patient engagement derived from the app were explored graphically and using descriptive statistics.

The primary efficacy end point, change in the total PANSS score from baseline to day 85 or last visit, was analyzed using the mixed effects model for repeated measures (MMRM) [25], applied on the intention-to-treat (ITT) set, which included all participants to whom study treatment was assigned by randomization and who had a baseline observation and at least one postrandomization observation for the analysis end point. The MMRM included fixed, categorical effects of treatment, visit, and treatment-by-visit interaction, as well as the continuous, fixed covariates of baseline score, baseline score–by–visit interaction, and disease duration at baseline. An unstructured covariance structure was used to model the within-patient errors. The Kenward-Roger method was used to adjust the estimated covariance of the mean difference and df. The primary comparison was the treatment contrast on day 85.

The secondary efficacy end points, including positive PANSS score, general psychopathology PANSS score, negative PANSS score, total MAP-SR score, total BDI-II score, and World Health Organization Quality of Life total scores for the four domains (social relationships, psychological, physical, and environment), were analyzed similarly to the primary end point (MMRM on change from baseline values, using the ITT set).

The retention to study treatment was assessed using Kaplan-Meier plots of time to drop out from any cause, on the *all randomized* set. Safety data and additional clinical outcome assessments were analyzed using descriptive statistics.

Ethics and Informed Consent

The study protocol was approved by the Copernicus Group Independent Review Board (study number 1251398). The study was conducted according to the International Conference on Harmonization E6 Guideline for Good Clinical Practice, which has its origin in the Declaration of Helsinki. Informed consent was obtained from each participant in writing at screening before any study-specific procedures were performed. The study was explained to the participant by the investigator or designee, who answered any questions, and written information was also provided.

Results

Study Sample

From December 10, 2018, to September 26, 2019, a total of 112 participants with schizophrenia were enrolled and randomized into the study (Figure 1). Of the 112 randomized participants, 92 (82.1%) completed the study (48/112, 85.7% from PEAR-004 and 44/112, 78.6% from the sham group). The most common reasons for discontinuation were lost to follow-up and participant or guardian decisions. One participant in the sham group discontinued because of an SAE of suicidal ideation. From the Kaplan-Meier plots (Figure 2), the observed time to discontinuation was somewhat longer in the PEAR-004 group than in the sham group; however, the difference was not statistically significant (log-rank test; P=.36).

Figure 2. Kaplan-Meier plot of time to dropout (all randomized participants).
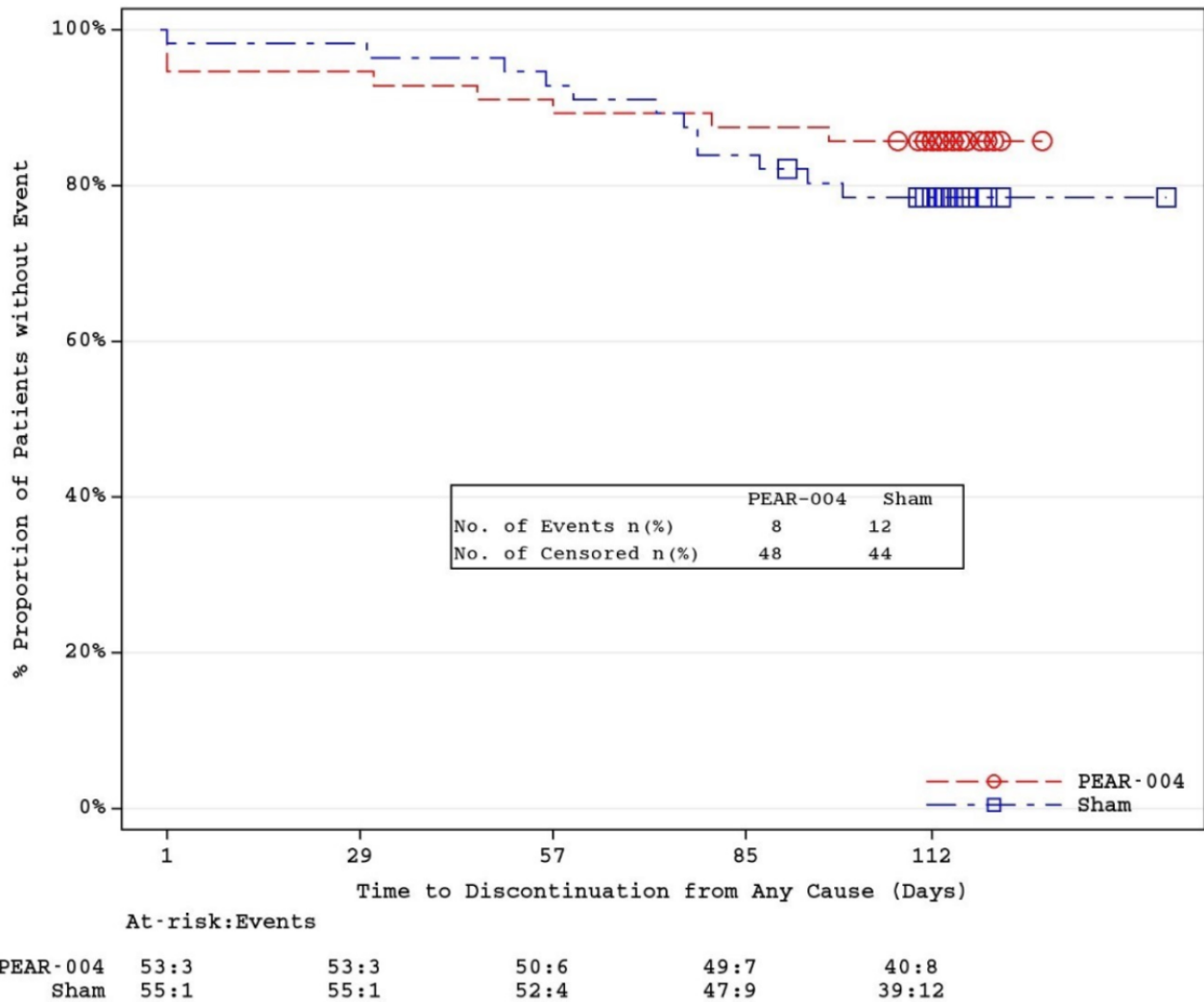
Demographics and Baseline Characteristics

Table 1 describes the demographics and baseline characteristics of study participants. Most participants were men (72/110, 65.5%) and Black or African American (53/110, 48.2%). The mean age of the participants was 45 (SD 11; range 22-65) years. Overall, the treatment groups were similar with respect to background disease characteristics. The most commonly rated participant's global severity was *moderately ill* (rating 4) in each treatment group (as rated on the CGI-Severity scale at baseline). The mean and median PANSS scores were comparable between the 2 randomized groups. The overall mean duration of disease before study entry was 17.2 (SD 11.5) years, which was balanced across the treatment groups.

Table 1. Participant demographics and baseline characteristics (safety analysis set).

Characteristic	PEAR-004 (n=55)	Sham (n=55)	Total (N=110)
Age (years)			
Value, mean (SD)	43.7 (10.99)	45.7 (11.60)	44.7 (11.29)
Value, median (range)	44.0 (24-64)	48.0 (22-65)	45.5 (22-65)
Sex, n (%)			
Male	36 (65.5)	36 (65.5)	72 (65.5)
Female	19 (34.5)	19 (34.5)	38 (34.5)
Race, n (%)			
American Indian or Alaska Native	3 (5.5)	0 (0)	3 (2.7)
Asian	3 (5.5)	4 (7.3)	7 (6.4)
Black or African American	28 (50.9)	25 (45.5)	53 (48.2)
Native Hawaiian or Other Pacific Islander	0 (0)	1 (1.8)	1 (0.9)
White	18 (32.7)	25 (45.5)	43 (39.1)
Other	3 (5.5)	0 (0)	3 (2.7)
Total PANSS[a] score at baseline			
Value, mean (SD)	73.5 (10.25)	72.7 (10.10)	73.1 (10.14)
Value, median (range)	72.0 (61-104)	71.0 (59-106)	72.0 (59-106)
CGI[b] severity at baseline, n (%)			
3	11 (20)	11 (20)	22 (20)
4	39 (70.9)	37 (67.3)	76 (69.1)
5	4 (7.3)	7 (12.7)	11 (10)
Missing	1 (1.8)	0 (0)	1 (0.9)
Disease duration at baseline (years)			
Value, mean (SD)	16.2 (11.40)	18.2 (11.56)	17.2 (11.47)
Value, median (range)	13.0 (1-46)	16.0 (1-47)	14.0 (1-47)

[a]PANSS: Positive and Negative Syndrome Scale.

[b]CGI: Clinical Global Impression.

Most participants were taking prior antipsychotic medication at the start of the study. Benzatropine, aripiprazole, quetiapine, and psychiatric medications were the most common medications used for the treatment of schizophrenia in both groups.

Engagement With App

Table 2 presents a summary of key engagement metrics by the treatment group. Participants in the PEAR-004 group spent significantly more time using the app over the course of the trial than participants in the sham group. However, the sham provided a good control for attention, as there were no significant differences between the groups in the number of days using the app, the total number of sessions, or the number of sessions per day.

Table 2. Engagement metrics for the PEAR-004 and sham groups.

| Variable	PEAR-004 (n=55)	Sham (n=55)	P value
Time (hours/day; weeks 1-12)			
Value, mean (SD)	4.2 (3.4)	2.2 (4.6)	<.001
Value, median (range)	3.4 (0-14)	0.8 (0-24)	N/A[a]
Number of days active			
Value, mean (SD)	62.1 (25.8)	63.7 (23.7)	.97
Value, median (range)	76 (3-87)	71 (1-86)	N/A
Number of sessions			
Value, mean (SD)	257 (160)	320 (295)	.36
Value, median (range)	241 (9-793)	278 (7-1805)	N/A
Number of sessions per day			
Value, mean (SD)	4.1 (1.7)	4.8 (3.2)	.15
Value, median (range)	3.6 (2-9)	3.9 (2-21)	N/A

[a]N/A: not applicable.

A more in-depth analysis of engagement data is provided in Tables S1 and S2 in Multimedia Appendix 1. Pear Therapeutics performed poststudy interviews with coordinators at all 6 study sites and found that there were differences among the sites' onboarding processes and directions on the product use at baseline and throughout the study. These differences in execution may have contributed to the statistically significant site differences in engagement measures observed in the PEAR-004 arm but not in the sham arm. Furthermore, we performed a visual exploration of the primary efficacy outcome (change in total PANSS score at day 85) for PEAR-004 and sham groups across 6 study sites (Figure S3 in Multimedia Appendix 1). There was no evidence that any given site had a between-group difference that would be inconsistent with the primary efficacy analysis of the pooled data across the sites (described in the *Efficacy* section). Therefore, site operational heterogeneity was associated with the measures of engagement with PEAR-004; however, this heterogeneity did not correlate with the observed difference in efficacy between the 2 groups.

Efficacy

In the primary efficacy end point of change in total PANSS score from baseline to day 85 or the last visit (Figure 3; Table 3), no benefit was seen with PEAR-004 versus sham. The estimated mean change in the PEAR-004 group was −1.6, −2.5, and −2.6 at days 29, 57, and 85, respectively. In the sham group, the estimated mean change was −2.2, −3.3, and −5.3 at days 29, 57, and 85, respectively. Thus, there was a small nondifferential improvement over time in both groups. At day 85, the treatment mean difference between PEAR-004 and the sham group was 2.7 points in favor of the sham (2-sided P=.09; 90% CI 0.1-5.4).

Figure 3. Mixed effects model for repeated measures—estimated mean change (SE) in the total Positive and Negative Syndrome Scale score over time for both groups. Chg: change. LSM: least squares mean. Ref: Reference.
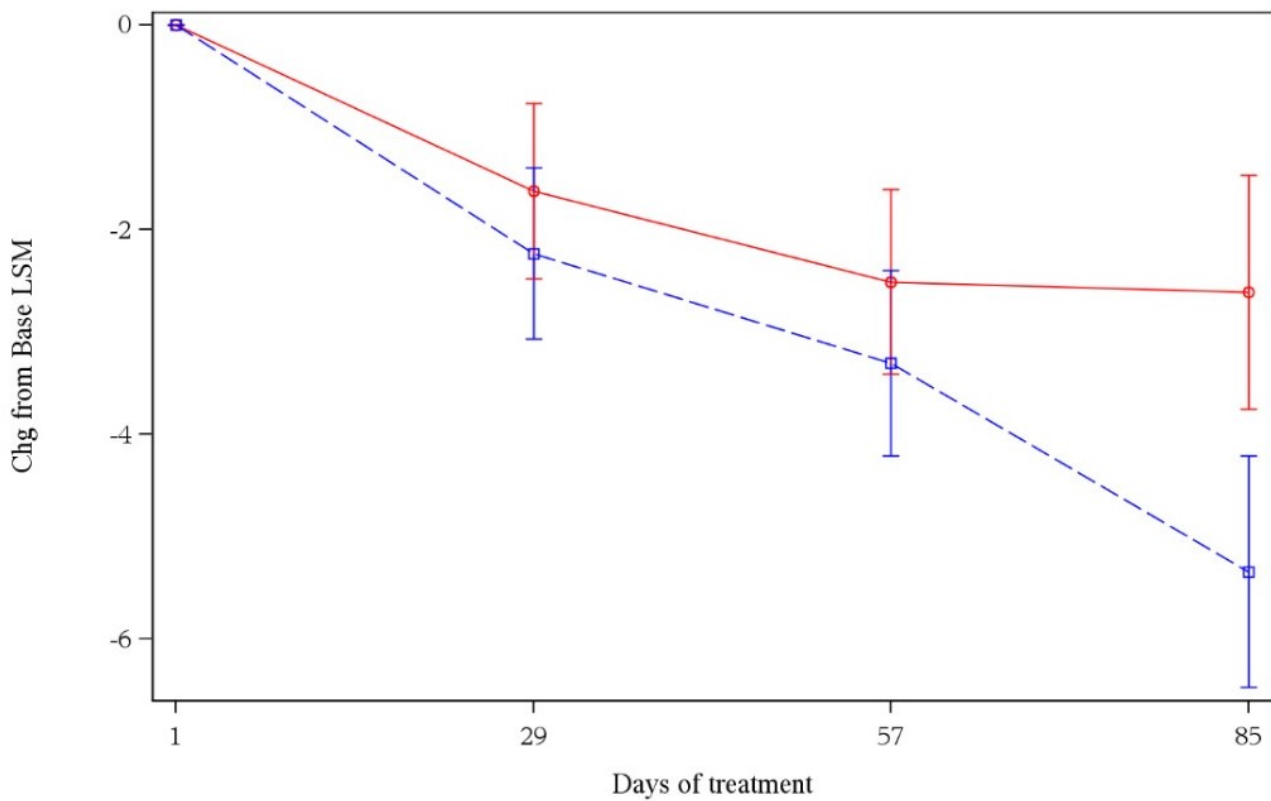
| Pear (Test) | . | 52 | 49 | 48 |
| sham (Ref.) | . | 54 | 47 | 49 |

Table 3. Mixed effects model for the repeated measures analysis of change in total Positive and Negative Syndrome Scale scores[a].

Test vs reference	Value, N			Adjusted least square means (SE)		Comparison of adjusted least square means (test vs reference)	
	Day	Test	Reference	Test	Reference	Difference, test–reference (SE; 90% CI)	P value
PEAR-004 (n=52) vs sham (n=54)							
Analysis 1	29	52	54	−1.6 (0.86)	−2.2 (0.84)	0.61 (1.200; −1.4 to 2.6)	N/A[b]
Analysis 2	57	49	47	−2.5 (0.9)	−3.3 (0.9)	0.80 (1.279; −1.3 to 2.9)	N/A
Analysis 3	85	48	49	−2.6 (1.14)	−5.3 (1.13)	2.73 (1.611; 0.1 to 5.4)	.09

[a]Baseline is the last measurement before treatment administration. Model: change from baseline of efficacy end point data was modeled using a mixed effects model with treatment, visit as fixed effects, baseline and disease duration at baseline as continuous covariates, treatment × visit, and baseline × visit interaction effects. The reported P value is 2-sided.

[b]N/A: not applicable.

Table 4 presents the results of secondary efficacy outcomes. No notable benefits were seen except for a small benefit for PEAR-004 in the BDI-II total score at day 57 (least squares mean difference of 3.3 points, PEAR-004 vs sham); however, this difference did not persist at day 85.

Table 4. Mixed effects model for the repeated measures analysis of change in secondary efficacy outcome measures[a].

Variable (PEAR-004 [N=52] vs sham [N=54])	Value, n			Adjusted least square means (SE)		Comparison of adjusted least square means (test vs reference)	
	Day	Test	Reference	Test	Reference	Difference, test–reference (SE; 90% CI)	P value
Positive PANSS[b]							
Analysis 1	29	52	54	–0.2 (0.36)	–0.4 (0.35)	0.21 (0.501; –0.6 to 1.0)	N/A[c]
Analysis 2	57	49	47	–0.8 (0.38)	–1.4 (0.38)	0.61 (0.538; –0.3 to 1.5)	N/A
Analysis 3	85	48	49	–1.0 (0.46)	–1.8 (0.45)	0.82 (0.648; –0.3 to 1.9)	.21
General psychopathology PANSS							
Analysis 1	29	52	54	–0.8 (0.61)	–1.4 (0.35)	0.51 (0.849; –0.9 to 1.9)	N/A
Analysis 2	57	49	47	–1.3 (0.63)	–1.5 (0.38)	0.12 (0.890; –1.4 to 1.6)	N/A
Analysis 3	85	48	49	–1.2 (0.76)	–2.8 (0.45)	1.61 (1.071; –0.2 to 3.4)	.14
Negative PANSS							
Analysis 1	29	52	54	–0.5 (0.31)	–0.4 (0.30)	–0.13 (0.430; –0.8 to 0.6)	N/A
Analysis 2	57	49	47	–0.3 (0.32)	–0.5 (0.32)	0.15 (0.453; –0.6 to 0.9)	N/A
Analysis 3	85	48	49	–0.4 (0.41)	–0.9 (0.41)	0.51 (0.581; –0.5 to 1.5)	.38
MAP-SR[d]							
Analysis 1	29	51	54	0.8 (1.05)	1.6 (1.02)	–0.79 (1.467; –3.2 to 1.6)	N/A
Analysis 2	57	49	47	–0.5 (1.41)	1.1 (1.43)	–1.60 (2.006; –4.9 to 1.7)	N/A
Analysis 3	85	48	49	–1.2 (1.26)	2.8 (1.25)	–4.07 (1.780; –7.0 to –1.1)	.02
BDI-II[e] total							
Analysis 1	29	52	54	–1.0 (1.13)	–0.1 (1.11)	–0.92 (–3.6 to 1.7; 1.587)	N/A
Analysis 2	57	49	47	–4.8 (1.14)	–1.5 (1.15)	–3.30 (1.629; –6.0 to –0.6)	N/A
Analysis 3	85	48	49	–3.4 (1.32)	–3.2 (1.3)	–0.19 (1.855; –3.3 to 2.9)	.92
WHOQOL-BREF[f] domain 1 total							
Analysis 1	29	52	54	–0.1 (0.37)	0.1 (0.36)	–0.18 (0.516; –1.0 to 0.7)	N/A
Analysis 2	57	49	47	–0.3 (0.44)	–0.3 (0.45)	0.02 (0.630; –1.0 to 1.1)	N/A
Analysis 3	85	48	49	0.2 (0.45)	0.1 (0.45)	0.02 (0.635; –1.0 to 1.1)	.98
WHOQOL-BREF domain 2 total							
Analysis 1	29	52	54	–0.6 (0.40)	–0.1 (0.39)	–0.46 (0.562; –1.4 to 0.5)	N/A
Analysis 2	57	49	47	–0.7 (0.37)	–0.6 (0.38)	–0.11 (0.528; –1.0 to 0.8)	N/A
Analysis 3	85	48	49	–0.5 (0.43)	0.1 (0.42)	–0.57 (0.603; –1.6 to 0.4)	.34
WHOQOL-BREF domain 3 total							
Analysis 1	29	52	54	–0.3 (0.28)	0.1 (0.27)	–0.36 (0.389; –1.0 to 0.3)	N/A
Analysis 2	57	49	47	–0.3 (0.33)	0.1 (0.33)	–0.33 (0.465; –1.1 to 0.4)	N/A
Analysis 3	85	48	49	0.5 (0.33)	0.3 (0.33)	0.19 (0.469; –0.6 to 1.0)	.68
WHOQOL-BREF domain 4 total							
Analysis 1	29	52	54	–0.2 (0.61)	–0.9 (0.60)	0.61 (0.853; –0.8 to 2.0)	N/A
Analysis 2	57	49	47	–0.5 (0.61)	–0.7 (0.61)	0.25 (0.866; –1.2 to 1.7)	N/A
Analysis 3	85	48	49	–0.1 (0.69)	1.1 (0.68)	–1.15 (0.968; –2.8 to 0.5)	.24

[a]Baseline is the last measurement before treatment administration. Model: change from baseline of efficacy end point data was modeled using a mixed effects model with treatment, visit as fixed effects, baseline and disease duration at baseline as continuous covariates, treatment × visit, and baseline × visit interaction effects. Reported P value is 2-sided.

XSL•FO
RenderX

Safety

AEs were reported in 20% (22/110) of the participants (Multimedia Appendix 2). The incidence of AEs was similar across both treatment groups: 22% (12/55) for PEAR-004 and 18% (10/55) for the sham group. All reported AEs were categorized as mild (20/110, 18.2%) or moderate (2/110, 1.8%) in severity. No severe AEs were reported. Most AEs were not suspected to be related to the treatment and resolved or were recovering at the end of the study. One SAE (suicidal ideation) was reported in the sham group, and the participant was discontinued from the study. This event was considered resolved on day 43, and was not suspected to be related to the treatment. No clinically significant abnormalities related to any of the vital signs were reported during the study.

Other Assessments

Table S4 in Multimedia Appendix 1 provides an assessment of the quality of blinding in the study using the sham control app. As shown, the sham app provided only a partial blinding effect, with the direction of bias of unblinding in favor of the PEAR-004 app. As the overall results showed a lack of benefit in the PEAR-004, any bias from partial unblinding would only make any interpretation of hidden potential benefit even more unlikely.

Multimedia Appendix 3 presents a summary of CGI scores at day 85 or the last visit. The proportions of participants with a score of 3 (mildly ill) or 4 (moderately ill) were similar at day 85 in the PEAR-004 group (42/48, 88%) and in the sham group (44/49, 90%). At day 85, 6 participants in the PEAR-004 group and 3 participants in the sham group were markedly ill (score=5). The percentage of participants with a global improvement rating of 4 (ie, no change) was somewhat higher in the PEAR-004 group (29/48, 60%) than in the sham group (23/49, 47%).

There was no change in mean sleep difficulties (ISI score) from baseline to the last visit in the PEAR-004 group (9.5, SD 7.34 vs 9.5, SD 6.82, respectively). A small numerical decrease (ie, improvement) in mean ISI scores was observed in the sham group (baseline: 11.2, SD 7.04; last visit: 9.8, SD 7.37, respectively).

Multimedia Appendix 4 presents a summary of the clinician and patient satisfaction surveys. The Clinician Satisfaction Survey included 4 questions on the clinician's experience with PEAR-004 (rated 1-7, with 1 indicating a highly negative response and 7 indicating a highly positive response), and the fifth question on how frequently they accessed the web dashboard during the study. Most of the clinicians' responses had scores 4, showing that the clinicians were satisfied with the PEAR-004 app. The dashboard was accessed at least once per month (21/49, 42% responses), and at least once per week (11/49, 22% responses).

The patient satisfaction survey included 8 questions on the participant's experience with PEAR-004, each rated from 1 to 7. Most responses to each question received a score of 6 or 7, suggesting that the majority found the app acceptable and usable. Most app use was during mid–late morning, 9 AM to noon (34/49, 69% respondents) and during late afternoon, 3 PM to 6 PM (25/49, 51% respondents). Most frequently, the app was used at home (46/49, 93% respondents) and at a public place (15/49, 30% respondents).

Discussion

Principal Findings

The primary objective of this study was to determine whether PEAR-004, a software-based intervention delivered via smartphone, can further reduce symptoms of schizophrenia as measured by the PANSS in participants, almost all of whom are currently on antipsychotic pharmacotherapy. No benefit was seen. Secondary outcomes suggested brief transient improvement in depressive symptoms only. No safety concerns were observed.

The lack of efficacy in this study was not because of a lack of engagement, which was demonstrated to be adequate, or lack of satisfaction. Despite the lack of definable clinical benefit, patients reported that the PEAR-004 app was engaging, interactive, and helped them feel better.

To our knowledge, this study is the largest RCT to date with a sham control group for schizophrenia. Lack of benefit when compared with sham may reflect the nonspecific aspects of benefit seen with digital interventions. It is notable that all participants improved, and if a waitlist control had been used rather than sham, we might have interpreted the results as suggesting mild benefit with the intervention. Natural history of recovery from an acute psychotic exacerbation is also relevant for improvement seen with 12 weeks of follow-up. The inclusion criteria of a moderate psychotic state (PANSS total score ≥60) might have allowed for lower levels of symptomatology, often associated with sham or placebo responses.

This study provides valuable insights that may be useful in future development programs for DTx. The development and selection of a sham control is an important design consideration. In our study, the sham app only included notifications 3 times per day, and when it was opened, it displayed a prescription timer for the remaining duration of app availability. However, this simple control intervention demonstrated a somewhat higher efficacy than PEAR-004. On the basis of feedback from poststudy interviews, the sham app was helpful in focusing attention away from internal stimuli (eg, hearing voices) to the present moment, similar to a mindfulness exercise. The notifications could have been an important ingredient too, as several participants in the poststudy interviews described the

receipt of notifications as being meaningful, given their reported lack of social connections. The sham may also present a lower barrier to engagement, which may have been helpful for participants experiencing a greater acuity of clinical symptoms.

Strengths and Limitations

This study had several strengths. This was a multicenter, randomized, sham-controlled, rater-blinded study design, in which study participants continued on their prescribed clinician-directed pharmacotherapy. Participants randomized to the PEAR-004 arm had access to the full clinical content and the logic of PEAR-004 app and of its therapeutic components. The sham control arm was chosen to account for the nonspecific effects of engagement with a smartphone. The sham app did not deliver any active coping skills (ingredients of a psychosocial intervention). The analysis of the primary and key secondary efficacy end points was performed on the ITT sample using the MMRM approach, considering all available data from the study participants. The amount of missing data was small, and the results were consistent across the different end points.

There was some heterogeneity in the study implementation aspects by different study sites. From poststudy interviews with coordinators at all 6 sites, we found that there were differences among the sites' onboarding processes and directions on product use at baseline and throughout the study. A post hoc analysis revealed some evidence of heterogeneity of engagement measures across study sites, which may be linked to a lack of consistency in how study interventions were delivered to participants (Multimedia Appendix 1). However, site operational heterogeneity did not correlate with the difference in efficacy between the 2 groups; in other words, there was no evidence that any site exhibited between-group differences that would be inconsistent with the primary efficacy analysis of the pooled data across the sites. A conclusion from this experience is that it is important to provide detailed instructions to study sites on operational aspects and ensure systematic and similar adherence to those processes.

Furthermore, although PEAR-004 offered clinical content derived from evidence-based treatments such as CBT, in-application support for practicing skills, and applying knowledge to daily life was very limited. This, when combined with the broad therapeutic focus on all patients with schizophrenia (instead of targeting a specific symptom or symptoms), may have inadvertently created an intervention that provided some help to many participants while not adequately supporting the specific needs of any particular participant.

There are several potential approaches for improving the content of PEAR-004 to provide a more personalized treatment delivery. First, to ensure adequate engagement with the CBT mechanism of action, the app could be designed with a utility of personalized goal setting and skill recommendations to help users make progress toward their goals. In addition, creating a web-based environment in which patients can receive just-in-time coaching and medical support may be beneficial. Second, foundational research to better understand users' needs and how they engage with the digital intervention is essential. The app development should be designed iteratively, evaluating both engagement data and clinical outcomes to see whether the

skills that are thought to be important for mechanism of action are adequately practiced; if not, the design should be improved and the assumptions should be re-evaluated in view of accrued experimental data. Finally, the primary goal of this study was to assess the effect of a CBT delivered as adjunct therapy through a mobile app. Additional digital ingredients to optimize patient outcomes, such as medication reminders, tracking of adherence to prescribed medication, and symptom tracking, could be implemented. A challenge would be to assess the added value of each ingredient (and possibly their combinations) when interpreting the trial results. Quantifying engagement with different components of the intervention and properly accounting for it in the analysis (eg, through regression modeling) may be worthwhile. A framework for developing and evaluating complex interventions [26] can be useful in guiding the clinical development of theory-based DTx interventions.

Comparison With Prior Work

Our results contrast with the three currently available RCTs of digital interventions in schizophrenia, of which all report benefits for delusions, hallucinations, or apathy or withdrawal. However, most of those studies (n=150) were computer-based, not smartphone-based [12]. The other two studies were small, with fewer than 45 participants in each study [10,11]. Further, 2 of the 3 studies had no sham control group and used only waitlist control [11,12]. Another important difference is that unlike previous digital interventions in schizophrenia, PEAR-004 was not integrated into clinical care and no personalized coaching was used to direct and support engagement with treatment. Directing and supporting engagement, whether by a clinician or a trained coach, can help tailor the therapeutic experience for a specific patient. As a prescription digital therapeutic, PEAR-004 was designed to be prescribed by a clinician to their patient and integrated into ongoing care. Not providing this support in this study meant that the burden of knowing what will help was placed on the patient, which may have affected how they engaged with the treatment.

It is instructive to compare the therapeutic content of PEAR-004 with that of some previous mobile-based digital health interventions for people with schizophrenia. Similar to FOCUS [13] and Actissist [10] apps, PEAR-004 content was organized into several domains of cognitive or behavioral strategies, and users received 3 daily notifications prompting them to engage with the app. However, the content of PEAR-004 also included some unique features, such as elements of gamification, to explicitly encourage and incentivize skill mastery through repeated skill practice for generalization of practice to daily life and promote lasting change. Such features may be worth exploring in future studies. In addition, the content of PEAR-004 and future DTx could potentially benefit from the inclusion of short-term and long-term goal settings tailored to specific needs of an individual and mechanisms to help the user achieve their goals. For instance, the PRIME intervention [11] had a self-identified goal setting in the cognitive and behavioral domains, and a supportive web-based community of both age-matched peers with schizophrenia spectrum disorders and motivational coaches.

In this study, the duration of treatment with PEAR-004 or sham was 12 weeks, which is consistent with the results of antipsychotic drug trials in schizophrenia to assess short-term benefits. Proof-of-concept studies of other investigational DTx in psychosis (eg, Actissist [10]; PRIME [11]) have also been conducted for 12 weeks in treatment duration. Furthermore, some other DTx products (eg, Food and Drug Administration–cleared reSET for substance use disorder [27] and reSET-O for opioid use disorder [28]) demonstrated evidence of efficacy following a 12-week treatment period. Although the *optimal* duration for a DTx intervention may vary across different indications, the 12-week treatment duration seems reasonable from the standpoint of balancing engagement and benefits of disease self-management. In this study, there was also a final follow-up visit at week 16. These data were examined descriptively, and the results are available upon reasonable request. Overall, no apparent between-group differences were observed with respect to primary or secondary efficacy outcome measures at week 16. As shown in Figure S5 in Multimedia Appendix 1, the changes in the total PANSS score at week 16 were very similar between the 2 groups.

Conclusions

In the largest-to-date randomized, sham-controlled study of a digital therapeutic in schizophrenia, PEAR-004 did not demonstrate an effect on the primary outcome of total PANSS scores compared with sham. The secondary and exploratory results also did not demonstrate any notable benefits, except for possible temporary improvement in depressive symptoms. Both clinical investigators and study patients provided high satisfaction ratings for the PEAR-004 app. The study provided many useful scientific and operational insights that can be used in further clinical development of PEAR-004 and other investigational DTx.

Conflicts of Interest

SNG, OS, and JVD were employed by and owned stock in Novartis.

Multimedia Appendix 1
Engagement with the PEAR-004 app by study site.
[DOCX File , 165 KB - formative_v6i3e29154_app1.docx]

Multimedia Appendix 2
Treatment-emergent adverse events by preferred term.
[DOCX File , 16 KB - formative_v6i3e29154_app2.docx]

Multimedia Appendix 3
Clinical Global Impression of Improvement at day 85 or last visit.
[DOCX File , 16 KB - formative_v6i3e29154_app3.docx]

Multimedia Appendix 4
Summary of the clinician and patient satisfaction surveys.
[DOCX File , 2246 KB - formative_v6i3e29154_app4.docx]

Multimedia Appendix 5
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 5431 KB - formative_v6i3e29154_app5.pdf]

References

1. Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. PLoS Med 2005 May;2(5):e141 [FREE Full text] [doi: 10.1371/journal.pmed.0020141] [Medline: 15916472]
2. Rector N, Beck A. Cognitive behavioral therapy for schizophrenia: an empirical review. J Nerv Ment Dis 2001 May;189(5):278-287. [doi: 10.1097/00005053-200105000-00002] [Medline: 11379970]
3. Cook SC, Schwartz AC, Kaslow NJ. Evidence-based psychotherapy: advantages and challenges. Neurotherapeutics 2017 Jul;14(3):537-545 [FREE Full text] [doi: 10.1007/s13311-017-0549-4] [Medline: 28653278]
4. Understanding DTx: a new category of medicine. Digital Therapeutics Alliance. URL: https://dtxalliance.org/understanding-dtx [accessed 2022-02-01]
5. Torous J, Luo J, Chan S. Mental health apps: what to tell patients. Current Psychiatry 2018;17(3):21-25 [FREE Full text]

XSL•FO
RenderX

6. Graham AK, Lattie EG, Mohr DC. Experimental therapeutics for digital mental health. JAMA Psychiatry 2019 Dec 01;76(12):1223-1224 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.2075] [Medline: 31433448]

7. Firth J, Torous J. Smartphone apps for schizophrenia: a systematic review. JMIR Mhealth Uhealth 2015 Nov 06;3(4):e102 [FREE Full text] [doi: 10.2196/mhealth.4930] [Medline: 26546039]

8. Gire N, Farooq S, Naeem F, Duxbury J, McKeown M, Kundi PS, et al. mHealth based interventions for the assessment and treatment of psychotic disorders: a systematic review. Mhealth 2017;3:33 [FREE Full text] [doi: 10.21037/mhealth.2017.07.03] [Medline: 28894743]

9. Sverdlov O, van Dam J, Hannesdottir K, Thornton-Wells T. Digital therapeutics: an integral component of digital innovation in drug development. Clin Pharmacol Ther 2018 Jul;104(1):72-80. [doi: 10.1002/cpt.1036] [Medline: 29377057]

10. Bucci S, Barrowclough C, Ainsworth J, Machin M, Morris R, Berry K, et al. Actissist: proof-of-concept trial of a theory-driven digital intervention for psychosis. Schizophr Bull 2018 Aug 20;44(5):1070-1080 [FREE Full text] [doi: 10.1093/schbul/sby032] [Medline: 29566206]

11. Schlosser D, Campellone T, Truong B, Etter K, Vergani S, Komaiko K, et al. Efficacy of PRIME, a mobile app intervention designed to improve motivation in young people with schizophrenia. Schizophr Bull 2018 Aug 20;44(5):1010-1020 [FREE Full text] [doi: 10.1093/schbul/sby078] [Medline: 29939367]

12. Craig TK, Rus-Calafell M, Ward T, Leff JP, Huckvale M, Howarth E, et al. AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. Lancet Psychiatry 2018 Jan;5(1):31-40. [doi: 10.1016/S2215-0366(17)30427-3]

13. Ben-Zeev D, Kaiser SM, Brenner CJ, Begale M, Duffecy J, Mohr DC. Development and usability testing of FOCUS: a smartphone system for self-management of schizophrenia. Psychiatr Rehabil J 2013 Dec;36(4):289-296 [FREE Full text] [doi: 10.1037/prj0000019] [Medline: 24015913]

14. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov;40(6):1244-1253 [FREE Full text] [doi: 10.1093/schbul/sbu033] [Medline: 24609454]

15. Ben-Zeev D, Scherer EA, Gottlieb JD, Rotondi AJ, Brunette MF, Achtyes ED, et al. mHealth for schizophrenia: patient engagement with a mobile phone intervention following hospital discharge. JMIR Ment Health 2016 Jul 27;3(3):e34 [FREE Full text] [doi: 10.2196/mental.6348] [Medline: 27465803]

16. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull 1987;13(2):261-276. [doi: 10.1093/schbul/13.2.261] [Medline: 3616518]

17. Llerena K, Park SG, McCarthy JM, Couture SM, Bennett ME, Blanchard JJ. The Motivation and Pleasure Scale-Self-Report (MAP-SR): reliability and validity of a self-report measure of negative symptoms. Compr Psychiatry 2013 Jul;54(5):568-574 [FREE Full text] [doi: 10.1016/j.comppsych.2012.12.001] [Medline: 23351831]

18. Beck A, Steer R, Brown G. Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation; 1996.

19. The World Health Organization Quality of Life (WHOQOL). World Health Organization. URL: https://www.who.int/substance_abuse/research_tools/en/english_whoqol.pdf [accessed 2022-02-01]

20. Svarstad BL, Chewning BA, Sleath BL, Claesson C. The brief medication questionnaire: a tool for screening patient adherence and barriers to adherence. Patient Educ Counseling 1999 Jun;37(2):113-124. [doi: 10.1016/s0738-3991(98)00107-4]

21. Ayer DW, Jayathilake K, Meltzer HY. The InterSePT suicide scale for prediction of imminent suicidal behaviors. Psychiatry Res 2008 Oct 30;161(1):87-96. [doi: 10.1016/j.psychres.2007.07.029] [Medline: 18786729]

22. Sheehan DV, Alphs LD, Mao L, Li Q, May RS, Bruer EH, et al. Comparative validation of the S-STS, the ISST-Plus, and the C-SSRS for assessing the suicidal thinking and behavior FDA 2012 suicidality categories. Innov Clin Neurosci 2014 Sep;11(9-10):32-46 [FREE Full text] [Medline: 25520887]

23. Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. Psychiatry (Edgmont) 2007 Jul;4(7):28-37 [FREE Full text] [Medline: 20526405]

24. Morin C, Belleville G, Bélanger L, Ivers H. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011 May 01;34(5):601-608 [FREE Full text] [doi: 10.1093/sleep/34.5.601] [Medline: 21532953]

25. Mallinckrodt CH, Clark WS, David SR. Accounting for dropout bias using mixed-effects models. J Biopharmaceutical Stat 2011 Jul 24;11(1-2):9-21. [doi: 10.1081/bip-100104194]

26. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]

27. De novo classification request for reset. De Novo Summary. URL: https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN160018.pdf?forcedefault=true [accessed 2022-02-01]

28. Computerized behavioral therapy device for psychiatric disorders. US Food and Drug Administration. URL: https://www.accessdata.fda.gov/cdrh_docs/pdf17/K173681.pdf [accessed 2022-02-01]

Abbreviations

AE: adverse event
AVATAR: Audio Visual Assisted Therapy Aid for Refractory auditory hallucinations
BDI-II: Beck Depression Inventory, Second Edition
CBT: cognitive behavioral therapy
CGI: Clinical Global Impression
CONSORT: Consolidated Standards of Reporting Trials
DTx: digital therapeutics
IRT: interactive response technology
ISI: Insomnia Severity Index
ITT: intention-to-treat
MAP-SR: Motivation and Pleasure-Self Report
MMRM: mixed effects model for repeated measures
PANSS: Positive and Negative Syndrome Scale
PI: principal investigator
PRIME: Personalized Real-Time Intervention for Motivation Enhancement
RCT: randomized controlled trial
SAE: serious adverse event

XSL•FO
RenderX

Original Paper

A Novel Experience Sampling Method Tool Integrating Momentary Assessments of Cognitive Biases: Two Compliance, Usability, and Measurement Reactivity Studies

Teresa Boemo[1], MD; Angela Socastro[1], MD; Ivan Blanco[2], PhD; Oscar Martin-Garcia[1], MD; Ana Mar Pacheco-Romero[1], MD; Raquel Rodríguez-Carvajal[2], PhD; Alvaro Sanchez-Lopez[1], PhD

[1]Complutense University of Madrid, Madrid, Spain

[2]Autonoma University of Madrid, Madrid, Spain

Corresponding Author:
Teresa Boemo, MD
Complutense University of Madrid
Campus de Somosaguas, s/n
Madrid, 28223
Spain
Phone: 34 674107873
Email: mboemo@ucm.es

Abstract

Background: Experience sampling methods (ESMs) are increasingly being used to study ecological emotion dynamics in daily functioning through repeated assessments taken over several days. However, most of these ESM approaches are only based on self-report assessments, and therefore, studies on the ecological trajectories of their underlying mechanisms are scarce (ie, cognitive biases) and require evaluation through experimental tasks. We developed a novel ESM tool that integrates self-report measures of emotion and emotion regulation with a previously validated app-based cognitive task that allows for the assessment of underlying mechanisms during daily functioning.

Objective: The objective of the study is to test this new tool and study its usability and the possible factors related to compliance with it in terms of latency and missing responses. Among the compliance predictors, we considered psychological and time-related variables, as well as usability, measurement reactivity, and participants' satisfaction with the tool.

Methods: We conducted 2 extensive ESM studies—study 1 (N=84; a total of 3 assessments per day for 5 days) and study 2 (N=135; a total of 3 assessments per day for 10 days).

Results: In both studies, participants found the tool highly usable (average usability score >81). By using mixed regression models, we found both common and specific results for the compliance predictors. In both study 1 and study 2, latency was significantly predicted by the day ($P<.001$ and $P=.003$, respectively). Participants showed slower responses to the notification as the days of the study progressed. In study 2 but not in study 1, latency was further predicted by individual differences in overload with the use of the app, and missing responses were accounted for by individual differences in stress reactivity to notifications ($P=.04$). Thus, by using a more extensive design, participants who experienced higher overload during the study were characterized by slower responses to notifications ($P=.01$), whereas those who experienced higher stress reactivity to the notification system were characterized by higher missing responses.

Conclusions: The new tool had high levels of usability. Furthermore, the study of compliance is of enormous importance when implementing novel ESM methods, including app-based cognitive tasks. The main predictors of latency and missing responses found across studies, specifically when using extensive ESM protocols (study 2), are methodology-related variables. Future research that integrates cognitive tasks in ESM designs should take these results into consideration by performing accurate estimations of participants' response rates to facilitate the optimal quality of novel eHealth approaches, as in this study.

KEYWORDS

experience sampling method; compliance; usability; measurement reactivity; emotion; cognitive biases

XSL•FO
RenderX

Introduction

Mood and anxiety disorders are configured by a series of dysfunctional cognitive, emotional, and behavioral factors that occur over specific periods and are inherently dynamic: (1) they are influenced (predicted) by other thoughts, emotions, and behaviors preceding them, and (2) they also have consequences in the subsequent cognitions, moods, and behaviors of the individual. Although this dynamic interplay among psychological processes is supported by empirical research [1], and it was originally formulated by conceptual models guiding current treatments [2], this view has typically been ignored by standard diagnostic approaches. Diagnostic criteria (eg, Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition [3]) are mainly focused on determining the existence of a given dysfunctional process for a given predefined period (eg, depressed mood during the past 2 weeks and generalized worry during the past 6 months), without considering their potential interplays with ongoing cognitive-affective processes that might be causing or maintaining the problems across time. The obvious consequence of ignoring time-varying data in this issue is a lack of knowledge on the proximal cognitive-affective mechanisms potentially driving the onset, maintenance, and recovery of affective dysfunctions [4].

In this line, to better understand the underlying mechanisms implicated in affective psychopathology, it is necessary to consider the ecological daily life dynamics of these cognitive-affective processes. Experience sampling methods (ESMs) and ecological momentary assessments [5] have emerged as crucial techniques to advance our knowledge on the dynamic psychological systems accounting for mental health and affective dysfunctions. They comprise repeated measures designs, where psychological assessments are performed several times a day for several days. The use of these methods has many advantages in addition to the rich and detailed information they provide. ESMs support in situ evaluations (state measures), which reduce the memory bias for self-reported retrospective assessments [6]. Furthermore, the clear improvement in the ecological validity of ESMs, in contrast to trait-based questionnaires, permits greater generalizability. It allows for the investigation of different individuals in their own contexts across time and situations, enriching theoretical and practical knowledge about the cognitive-affective processes of mental health and well-being [7]. To date, ESM research has clearly advanced the understanding of psychological processes involved in mental health, such as affective emotional reactivity [8], emotion regulation dynamics [9], the ecological use of simultaneous emotion regulation strategies [10], and the specific effectiveness of these strategies to regulate different momentary affective states [11]. However, these ESM measurements are still solely based on self-reports and do not allow for the evaluation of the underlying processes of these affective dynamics, namely cognitive-affective mechanisms. Individual preferences in the way that emotional information is attended to or interpreted (ie, cognitive biases) are typically assessed using experimental technologies such as eye tracking [12] in controlled laboratory conditions. These experimental methods allow for the capture of subtle mechanisms related to affective

disorders [13] and the differential use of emotion regulation strategies [14-16] or even individual differences in psychological well-being [17]. However, to date, no research has fully integrated this type of cognitive bias assessment into ESM protocols, thus lacking a proper understanding of how their ecological momentary manifestations affect emotional experiences and their regulation in daily life.

Therefore, we developed a highly novel app system integrating a novel app-based cognitive task that allows for the measurement of attention and interpretation biases into a new ESM tool. This new method combines self-reported assessments of mood, stress appraisals, and emotion regulation use with momentary assessments of attention and interpretation biases during real-life functioning. Such assessments are based on the computerized version of the Scrambled Sentence Task (SST) [17], where participants are asked to freely unscramble a series of 6 scrambled word sentences displayed on the screen (eg, *born loser am I winner a*) using only 5 out of 6 words. The only 2 possible solutions to resolve the sentence are into a positive or a negative meaning (eg, *I am a born winner* or *I am a born loser*, respectively), and attentional processing of positive (eg, *winner*) and negative (eg, *loser*) are further assessed through advanced eye-tracking–based techniques [18]. In a series of experimental studies using this task, it was found that participants showing higher biases toward negative over positive information have poorer abilities to use emotion regulation strategies supporting negative affect downregulation (lower use of reappraisal and higher use of rumination) [19]. Furthermore, previous studies using the presented novel web-based app have demonstrated that ecological assessments of these negative cognitive biases assessed with this app-based cognitive task through mobile phones are predictive of poorer abilities to use emotion regulation strategies in their daily life functioning, ultimately leading to increased depressive and anxiety symptoms and reduced well-being in the face of major stressors [20]. Thus, given the large potential of these novel technical approaches to inform advanced health-related research and technology innovations, a thorough analysis of their usability and the conditions that facilitate their compliance is required. Ultimately, the aim of this study is to test the usability and feasibility of our method to be implemented for advanced research on the ecological mechanisms of mental health. Specifically, we first aim to study users' perceived usability and satisfaction with the novel system, which integrates cognitive tasks within extensive ESM procedures. Second, we aim to establish the degree of measurement reactivity to the new method. Finally, we aim to establish factors that must be considered to maximize compliance with the use of the method, both in terms of latency and missing responses. For the latter, we exhaustively examine what factors are involved in effective compliance with this type of new ESM system, considering usability, satisfaction, reactivity, and emotional symptoms, as well as time-related variables as predictors.

Potential factors accounting for differences in compliance with the new ESM system were pre-established based on previously identified factors involved in the compliant use of ESM procedures in general [5,21,22]. Compliance with ESM protocols can be defined in two ways: (1) effectively answering

the scheduled signals at the arranged times and (2) preventing the missing of assessment notifications. Therefore, it comprises time latencies and missing responses. Time latency is referred to as the time lag participants require to respond to a prompt or signal to start completing a momentary assessment, which ranges across studies from 90 seconds [23] to 3 hours [24]. Studying the latency of responses has several implications for the methodological quality of ESM studies and the quantity of available data, and it is closely related to missing rates.

Dealing with low compliance and abandonment rates has become a significant challenge when applying ESM designs in general. Owing to this, there is growing evidence exploring the possible systematic predictors of larger response latencies and missing data.

First, in terms of time-related variables, several studies have consistently found that missing rates tend to increase as ESM studies extend over time [25-30], although with some exceptions [31] (see available meta-analyses [29,32,33]). Furthermore, the effects of time of day (whether responding to ESM prompt signals during the morning, afternoon, or evening) are consistently related to missing responses [30,34,35]. For instance, some studies have found higher compliance rates in the afternoon (between 12 PM and 1:30 PM) and lower compliance rates in the morning (between 7:30 AM and 9 AM) [28]. Hence, in line with the revised literature, it is hypothesized that lower compliance with the new ESM system (increases in latency and missing responses) would emerge as the days of the study progress, as well as for earlier notifications during the day.

Second, there are relevant individual-related variables that should be considered when applying ESM procedures in general and specifically in ESM systems assessing mechanisms of mental health, such as the one proposed in this research. This refers to the consideration of how participants' psychopathological conditions affect their compliance with ESM. This issue has been considered in previous research and has shown different results. Some studies have not found differences in ESM compliance among clinical conditions (schizophrenia, substance dependence, and anxiety disorders vs no diagnosis) [31]. However, other studies have found that higher missing rates are predicted by various clinical conditions, including anxiety and depression levels [29,30,32-35]. Therefore, the second aim of this study is to test whether individual differences in emotional symptoms (depression and anxiety levels) affect compliance with ESM. Considering that in this study, we evaluate 2 samples with subclinical symptom levels, it is hypothesized that no associations would be found between individual levels of depression and anxiety and compliance rates, supporting the feasibility of the ESM system for its use in this type of population.

Third, we considered it crucial to explore the implications that ESM approaches such as the one proposed have on individuals' momentary experiences. Responding to ESM assessments at various times during the day entails paying regular attention to internal states and behaviors. Different results have been found regarding how the frequency of daily ESM assessments affects the dynamic trajectory of the psychological phenomenon being measured, thus generating different forms of measurement reactivity [36-38]. To date, few studies have evaluated the relationship between ESM compliance rates and experienced negative (burden) or positive (usefulness) measurement reactivity ESM. A recent study found an association between perceived burden and higher missing rates in longer ESM protocols [23]. According to this, to test the new ESM system developed for this study, we collected information regarding several measurement reactivity indicators that could be associated with different levels of compliance: experienced overload and stress generated by evaluation requirements, perceived usefulness of the assessment, and general satisfaction with the procedure. However, given the novelty of the topic under study, only the hypothesis regarding the relationship between burden and compliance is described. It is hypothesized that experiencing stress and overload would be related to lower compliance with the ESM protocol [23]. We further explore the relationship between positive reactivity variables, such as satisfaction and usefulness, and compliance rates during ESM because of the lack of previous literature on this topic.

Finally, when testing a new technological tool such as the one developed in this study, it is necessary to establish its degree of usability in terms of the system's ease of use and learnability. Several studies have reported that patients with psychological disorders often find difficulties in engaging with technology that is challenging to use or that is perceived as irrelevant to their needs [39,40]. Therefore, given the purpose of this new ESM tool (ie, ultimately implementing it for use with clinical samples), we aim to initially test it in subclinical samples, considering the usability of the tool, analyzing its relationship with individual differences in depression and anxiety levels, and testing whether different individual levels of perceived usability have an influence on compliance with the novel ESM protocol. There is no previous evidence on this topic, despite its clear importance. Owing to this fact, no specific hypotheses were made about the relationship between usability and participants' compliance with the new ESM approach.

To test these research issues, we conducted two studies testing the new app-based tool with different ESM regimes: The first study comprised a controlled experiment of 1 week (ie, study 1: a total of 5 days), and the second study extended the used ESM system to a larger period, comparable with regimes used in other types of ESM studies (ie, study 2: a total of 10 days), as described in the following sections.

Study 1

Methods

Participants

A sample of 84 undergraduate students (age: mean 20.05, SD 2.19 years) was recruited from the Faculties of Psychology of Complutense University of Madrid and the Autonomous University of Madrid. The participants received extra credit for participating in the study.

Procedure

Participants attended an introduction session in groups of 15 to 20, during which they received information about the study protocol and completed the informed consent forms, as well as baseline questionnaires assessing demographic variables and depressive and anxiety symptom levels. During the introductory session, participants downloaded the ESM tool, comprising a mobile app, and performed practice trials to become familiar with its use. A day after the introduction session, participants were instructed to complete the ESM assessment each time they received a new signal notification. A systematic sampling approach was used to determine random signaling schedules. Experience sampling assessments were programmed to be sent to participants 3 times a day for 5 days. These assessments were prompted randomly between 9 AM and 9 PM at three time intervals (9 AM to 1 PM, 1 PM to 5 PM, and 5 PM to 9 PM). Participants had 1 hour since they received the notification to complete the assessment. In each ESM assessment, they completed measures of stress, current affect, and use of emotion regulation strategies and performed a cognitive bias task. Furthermore, using the software, we generated a database where we gathered compliance-related information, such as the latency of response, missing assessments, and abandonment rates. At the end of the study, participants completed a brief questionnaire that accounted for variables related to measurement reactivity (app stress, app overload, and app usefulness) and user experience (usability and satisfaction). Questionnaires at baseline and the end of the study were gathered using Qualtrics (Qualtrics International Inc) software.

Instruments

Baseline Measures

In the initial introduction session, participants rated their levels of depression and anxiety, which were measured with the Center for Epidemiologic Studies–Depression (CES-D 8 [41]) scale and the Generalized Anxiety Disorder-7 (GAD-7 [42]) scale, respectively.

The CES-D 8 [41] is a screening scale used to evaluate depressive symptom levels in the past week. It contains 8 items that evaluate the severity of symptoms, which show good reliability in both general and clinically depressed samples [43,44]. The internal consistency in this study was α=.86.

GAD-7 [42] is a 7-item screening questionnaire that evaluates the severity of anxiety symptom levels comprising emotional and cognitive symptoms of anxiety. This measure has good reliability and validity in both general and clinically anxious samples [45,46]. The internal consistency in this study was α=.89.

ESM Assessments

The app comprised momentary self-reported assessments of several psychological states, including self-reports of perceived ongoing stress, use of emotion regulation strategies in response to ongoing stress, motivational factors, and current mood states.

Participants also completed a novel app-based cognitive task assessing momentary attention and interpretation biases in each survey. We created a computerized SST [17] that allows for app-based assessments of these cognitive biases. This task has been previously validated for the evaluation of attention and interpretation bias in various formats, such as computers [18], and it has already been used in mobile phones for implementation in ESM procedures [20]. Participants were required to complete 20 sentences at each ESM assessment a total of 3 times a day for 5 days. The full details of the ESM assessments in the new tool are provided in Multimedia Appendix 1.

Assessments at the End of the Study

Once participants completed the ESM part of the study, they received a final questionnaire to measure variables related to the use of the app: usability, satisfaction, app stress, app overload, and app usefulness. They completed these questionnaires using Qualtrics. To date, there are no validated evaluation protocols to assess these characteristics for eHealth mobile apps. Therefore, for some of the variables of interest, we opted to select single items from the Mobile Application Rating Scale [47] that best resembled the following measurement reactivity variables of interest: app satisfaction, app stress, app overload, and app usefulness.

App stress was measured to control for the reactivity of participants while performing ESM with this app with the item "I felt more nervous than usual, while being vigilant to receive the App's notifications."

App overload was used to control for whether the length of the ESM assessments (ie, self-reported surveys and trials of the app-based cognitive task) was generating overload in participants: "The number of exercises to perform in each signal was excessive."

We also included an item to measure whether participants found the app useful in terms of facilitated introspection to understand their feelings, cognitions, and behaviors (app usefulness) with the item "Using the App has helped me to be more conscious about my emotional and cognitive responses through the day and across days."

Global satisfaction with the app was assessed using a Likert scale ranging from 0 to 10, with 10 being highly satisfied: "Indicate, from 0 to 10, your overall satisfaction with the app."

Finally, we used the System Usability Scale (SUS) [48] to measure the usability of the new ESM tool. Usability was assessed at the end of the study to estimate whether the new software had been experienced as usable. This scale has been previously validated [49], and it measures aspects such as complexity, technical support, integration, consistency, and general satisfaction. We computed the overall score following the indications of the author, which ranged from 0 to 100, with scores >68 being considered above average.

App Compliance

Compliance was defined in terms of two different dependent variables: latency and missing responses. Latency was indexed by the time participants took to respond to each signal (ie, the time lag between the prompt and the actual response, which had a maximum of 60 minutes). Missing responses were indexed

using scheduled ESM prompts that were not completed by the participant.

Analytic Plan

We first conducted descriptive analyses of the data and used correlation analyses to study the relationships between participants' symptoms (depression and anxiety levels), usability scores, and variables related to measurement reactivity (app stress, app overload, app usefulness, and satisfaction).

Then, to test factors accounting for individual differences in compliance, we ran multilevel analyses taking into account the nested structure of the data because of the repeated measures design (ie, observations nested within days and days nested within persons). This permitted the examination of the momentary variation of compliance variables (latency and missing responses) across prompts, considering the variability in intrapersonal and interpersonal variables. We used the *lme4* R package [50] to conduct the models predicting latency responses and missing responses. We applied the function *glmer* for latency because of the reaction time characteristics [51], specifying the family as Gamma. The *glmer* function was also used in models predicting missing responses, as this variable was coded as categorical, specifying family as a binomial. In all the models, we specified crossed random effects at the individual level. All models were fit by maximum likelihood estimation.

We first modeled an empty model, with each compliance variable predicted by its intercept. After that, we added one predictor variable at a time (univariate models) and then fitted a model with all predictors included simultaneously (multivariable model) to test whether the effects of predictor variables changed once the remaining variables were included. Fixed slopes were specified for all the models.

Thus, the time of day was entered as a level 1 predictor and day as a level 2 predictor. To explore whether individual differences in depression and anxiety levels affected compliance during the ESM study, these variables were introduced as level 3 predictors.

Variables such as app stress, app overload, app usefulness, global satisfaction with the app, and usability were also introduced as level 3 predictors. We performed grand centering transformation of all the variables introduced as level 3 predictors.

Ethics Approval

The study was approved by the Faculty Ethical Committee of Complutense University of Madrid (Protocol Code Ref. 2019/20-028) and complied with the Declaration of Helsinki's ethics standards.

Results

Sample Characteristics and General App Performance

Initially, we recruited 102 university students. The level of abandonment was low; only 8.8% (9/102) of participants stopped responding to the ESMs and did not finish the study. Of those 9 participants, 5 (56%) missed the last 5 ESM prompts, and 4 (44%) responded only once. In addition, 8.8% (9/102) participants did not respond to the last questionnaire after completing the ESM protocol. Therefore, of the 102 participants, the final sample comprised 84 (82.4%) participants, with a mean age of 20.03 (SD 2.19) years, ranging from 18 to 29 years. We found a low mean number of missing responses per participant—1.43 (SD 1.97), ranging from 0 to 8 missing responses per participant. The mean levels of latency per participant found in this study were 16.38 (SD 7.37), ranging from a minimum of 3.24 minutes to a maximum of 34.37 minutes per participant. As shown in Table 1, participants presented mild levels of depression and anxiety, as measured with the CES-D and GAD-7, respectively. Mean scores for measurement reactivity showed moderately high levels of stress related to the use of the app. The mean app overload showed medium levels, pointing that the number of measures per assessment (questionnaires and cognitive bias tasks) was not burdensome. The extent to which participants found the app useful reflected moderately high scores, whereas their general satisfaction with the app showed medium levels.

Table 1. Sample characteristics and general app performance (study 1; N=84).

Characteristics	Study 1 sample
Sex (female), n (%)	75 (89)
Age (years), mean (SD)	20.03 (2.19)
Anxiety (GAD-7[a]; 0-21), mean (SD)	5.55 (3.82)
Depression (CES-D[b] 8; 0-24), mean (SD)	7.64 (4.07)
App stress (1-5), mean (SD)	3.38 (1.50)
App overload (1-5), mean (SD)	2.69 (1,24)
App usefulness (1-5), mean (SD)	3.18 (1.30)
Satisfaction with the app (1-10), mean (SD)	6.95 (1.61)
Usability (SUS[c]; 0-100), mean (SD)	81.01 (12.08)

[a]GAD-7: Generalized Anxiety Disorder-7.

[b]CES-D: Center for Epidemiologic Studies–Depression.

[c]SUS: System Usability Scale.

Importantly, participants reported high levels of usability measured by the SUS.

Out of 100, it reached a mean usability score of 81.01 (SD 12.08), reflecting its ease of use and learnability. According to the authors who validated the scale [49], scores >70 are considered above average and acceptable, and >80.30 is in 10% of the best-rated systems [49].

Correlation Analyses

Correlation analyses were conducted to test how the levels of emotional psychopathology (ie, depression and anxiety levels) are related to measurement reactivity and usability indices. Given the nonnormal distribution of these variables, we conducted Spearman correlation analyses, which are shown in Table S1 in Multimedia Appendix 1. First, we found a positive, significant relationship between depression and anxiety levels ($r=0.67$; $P<.001$), indicating a relatively high degree of comorbidity among both types of symptoms. Nonetheless, neither anxiety nor depression levels were related to individual differences in the variables of user experience with the app or its usability, suggesting that such ratings were not affected by individual differences in symptom levels. Furthermore, we found significant correlations between usability and most of the measurement reactivity variables. Usability was positively related to app usefulness and satisfaction ($r=0.31$, $P=.004$ and $r=0.60$, $P<.001$, respectively) and negatively related to app overload ($r=-0.46$; $P<.001$) but not to app stress ($r=-0.21$; $P=.05$). These results show the importance of focusing on user experience when developing new eHealth app-based assessment methods, as it seems to be closely related to measurement reactivity and, in turn, the methodological quality of the design. In addition, app overload was significantly positively related to app stress ($r=0.23$; $P=.02$) and negatively related to satisfaction with the app ($r=-0.30$; $P=.008$), indicating that feeling burdensome because of the length of evaluations is related to the stress generated by the notifications' requirements and to lower satisfaction with the app.

Multilevel Analyses

Latency

We conducted a series of transformations to control for the distribution of the outliers. Outliers were substituted with the upper or lower threshold of each participant based on the IQR. After that, we calculated the interclass correlation coefficient (ICC) for the empty model, showing a value of 14% for between-person variance. After that, we performed linear mixed models to test the predictors of the variability of latency when responding to experience sampling.

First, we conducted a series of univariate models, including 1 predictor at a time, and then tested them in a multivariate model to determine whether those effects remained significant after the inclusion of all predictor variables simultaneously. This information can be found in Table 2.

Analyses testing the effects of time-related variables on latency showed a significant effect of the day on responses' latency in the univariate model (estimate 1.03 [SE 0.32]; $P<.001$), which remained significant in the multivariable model (estimate 1.00 [SE 0.22]; $P<.001$), indicating that a change of 1 unit on the score scale of the day (ie, 1 day passed) produced a 1.00 point of increase in latency. This indicated that as the study progressed, response latencies were longer. Furthermore, we did not find a significant effect of time of the day, neither in the univariate model nor in the multivariate model. Analyses testing the effects of emotional symptomatology on latency showed no significant effects of depression or anxiety levels when introduced as single predictors or in the multivariate model. Thus, participants' levels of symptomatology did not influence their response latencies. Analyses testing the relationship between app-related variables and latency rates showed no significant effects of any of the variables when introduced as single predictors in the univariate models. In the multivariate model, a significant effect of app usefulness was found (estimate 1.38 [SE 0.61]; $P=.02$), indicating that participants who found the app as more useful took more time to respond to the notifications (an increase of 1 point in the score scale of app usefulness was related with an increase of 1.38 points in latency).

Table 2. Predictors of response latency (study 1).

Outcome variable: latency	Estimate (SE)	*t* value	*P* value
Empty model			
Intercept	15.91 (0.74)	21.42	<.001
Univariate models			
Day	1.03 (0.22)	4.70	<.001
Time of day	−0.60 (0.37)	−1.61	.11
Depression	0.11 (0.19)	0.59	.55
Anxiety	−0.05 (0.20)	−0.26	.80
App stress	−0.63 (0.50)	−1.26	.21
App overload	0.14 (0.61)	0.24	.81
App usefulness	0.96 (0.56)	1.72	.09
App satisfaction	−0.30 (0.46)	−0.64	.52
Usability	0.02 (0.06)	0.41	.69
Multivariate model			
Day	0.10 (0.22)	4.58	<.001
Time of day	−0.53 (0.36)	−1.46	.15
Anxiety	−0.36 (0.27)	−1.31	.19
Depression	0.40 (0.26)	1.55	.12
App stress	−0.43 (0.50)	−0.86	.39
App overload	0.27 (0.64)	0.42	.68
App usefulness	1.38 (0.61)	2.28	.02
Satisfaction	−1.09 (0.60)	−1.81	.07
Usability	0.05 (0.08)	0.60	.55

Missing

The ICC calculation of the empty model predicting missing responses showed a value of 41% for between-person variance. Analyses testing the effect of time-related variables on missing responses showed significant effects of the day number both in the univariate model (estimate 0.23 [SE 0.08]; *P*=.004) and in the multivariable model (estimate 0.24 [SE 0.08]; *P*=.003). Time of the day did not show a significant effect on missing responses in the univariate or multivariate models. This information can be found in Table 3. Therefore, as the study progressed, the number of missing responses became higher, whereas the time of the day in which notifications were sent did not show an effect on missing responses. Analyses testing the effects of emotional symptomatology on missing responses showed no significant effects of depression or anxiety levels when introduced as single predictors or when included in the multivariate model. Finally, analyses testing the relationship between app-related variables and missing response rates showed no significant effects of app stress, app usefulness, app overload, satisfaction, or usability on missing responses. These effects were not significant in either the univariate or the multivariate models. Therefore, we can assume that technical, personal, or usability-related variables were not related to the differential missing response rates.

Table 3. Predictors of missing responses (study 1).

Outcome variable: missing	Estimate (SE)	z value	P value
Empty model			
Intercept	−3.14 (0.29)	−10.7	<.001
Univariate models			
Day	0.23 (0.08)	2.91	.004
Time of day	0.15 (0.13)	1.16	.25
Anxiety	0.01 (0.06)	0.2	.84
Depression	0.02 (0.05)	0.37	.71
App stress	−0.22 (0.15)	−1.51	.13
App overload	−0.27 (0.17)	−1.55	.12
App usefulness	−0.13 (0.17)	−0.81	.42
App satisfaction	−0.09 (0.14)	−0.68	.50
Usability	0.02 (0.02)	0.91	.37
Multivariate model			
Day	0.24 (0.08)	3.06	.002
Time of day	0.18 (0.13)	1.33	.18
Anxiety	0.03 (0.08)	0.37	.71
Depression	−0.01 (0.08)	−0.15	.88
App stress	−0.19 (0.15)	−1.26	.21
App overload	−0.28 (0.20)	−1.38	.17
App usefulness	−0.09 (0.19)	−0.50	.64
App satisfaction	−0.31 (0.19)	−1.62	.11
Usability	0.03 (0.03)	1.12	.26

Discussion (Study 1)

The main aim of this study was to examine the usability and feasibility (ie, compliant use) of a new ESM tool that integrates self-reported assessments of affective experience and a cognitive task assessing attention and interpretation biases. A series of multiple time, person, and system reactivity variables were tested as potential predictors of compliance. First, participants reported high levels of usability for the new ESM system, with scores on 10% of the best-rated systems, according to the SUS criteria. Compliance was also high in terms of both low latencies and missing responses. As for compliance predictors, we found significant effects of the day number on response latency and missing responses, indicating that latencies became longer, and there were higher missing rates as the study progressed. These results are in line with previous literature, which supports that lower compliance is found as the days of the study progress [35].

Individual differences in depression and anxiety levels did not act as significant predictors of compliance for either latency or missing responses. These results suggest that adequate compliance with the ESM system is not affected by participants' subclinical depression and anxiety levels, which is a crucial aspect of the aim of this type of new ESM system. To advance future clinical implementation, further research should analyze these issues in participants presenting with higher levels of

depression and anxiety symptomatology to test the effect of clinical status on compliance with this ESM system integrating both self-reports and cognitive tasks [36].

Furthermore, system-related variables were not significantly associated with compliance rates, except app usefulness. Interestingly, participants who found the app to be more useful were those who were slower in responding to momentary notifications. This might be indicative of an attempt to find the more proper moments to adequately perform the app tests and assessments within the allowed 1-hour period after notification. Nonetheless, this effect was only evident in the multivariate model, whereas the univariate analyses did not show any direct association and should be considered cautiously until further replication.

Overall, the results showed high usability and feasibility for the use of the new ESM system, with few factors substantially accounting for its adequate compliance. Nonetheless, we should note various limitations in this study. First, we used single items to evaluate variables related to measurement reactivity, such as app stress, overload, usefulness, and satisfaction, which might at least partly obscure some of their potential associations with compliance.

Furthermore, there might be certain limitations in the generalizability of the results, given the relatively short number

of days evaluated (ie, 5 days). These current findings invite further replication using the new ESM system in protocols with more extensive durations that are comparable with other ESM research considering compliance predictors. For these reasons, we performed a second ESM study with a longer protocol (ie, 10 days of assessments), integrating the study of measurement reactivity variables not only through single items but also through different validated scales.

Thus, study 2 overcame the abovementioned issues, adding further contributions to knowledge in particular areas. First, in study 2, we expanded the sample of participants and extended the duration of the study. It is important to verify whether the results, in terms of reactivity with the app and compliance, would change when participants were evaluated twice as long. In addition, we performed a significant methodological improvement by including validated subscales to measure the variables of app-related stress, overload, usefulness, and satisfaction to derive precise knowledge of their relationship with compliance rates.

Study 2

Methods

Participants

A sample of 135 undergraduate students (age: mean 20.52, SD 2.31 years) was recruited from the Faculty of Psychology at Complutense University of Madrid between April and May 2021. The participants received extra credit for participating in the study.

Procedure

Participants individually attended an app-based introductory session in which they received information about the study. They were trained to use the app in which the ESM system was integrated and performed a practice exercise of self-report measures and cognitive tasks. On the first day, they also completed the informed consent and a baseline questionnaire assessing demographic variables, measures of depression and anxiety levels, and other scales not relevant to the aim of this study. One day after the introductory session, participants were instructed to start completing the ESM protocol through the app on their phones each time they received a survey notification. A systematic sampling approach was used to determine the random signaling schedules. Experience sampling assessments were programmed to be sent to participants 3 times a day for 10 days. These assessments were prompted randomly between 10 AM and 9 PM at three time intervals (10 AM to 11 AM, 3 PM to 4 PM, and 8 PM to 9 PM). Participants had 1 hour since the time they received the notification to complete the assessment. At each assessment, they completed assessments of stress, current affect, and emotion regulation strategies and performed the cognitive bias task exactly as in study 1. Furthermore, as in study 1, compliance-related information, such as latency of response and missing and abandonment rates, was gathered. At the end of the study, participants completed a brief questionnaire that accounted for variables related to measurement reactivity and user experience. Questionnaires were completed at baseline and at the end of the study using Qualtrics software.

Instruments

Baseline Assessments

Depression and anxiety symptom levels were measured using the CES-D-8 [41] and the GAD-7 [42], respectively, as in study 1. The internal consistencies for the CES-D 8 and GAD-7 in this study were α=.88 and α=.86, respectively.

Momentary Assessments

As in study 1, assessments referring to self-reports of perceived ongoing stress, use of emotion regulation strategies, and current mood states were evaluated through app-based self-reports. In each signal, participants also completed a cognitive task assessing momentary attention and interpretation biases, which was based on the SST [17]. They were required to complete 15 phrases at each beep for a total of 3 times a day for 10 days. Further details of all assessments are provided in Multimedia Appendix 1.

Assessments at the End of the Study

As in study 1, after completing the ESM procedure, participants received the last questionnaire survey, which measured system usability, measurement reactivity, and user experience using the same assessments. Furthermore, study 2 also included the assessment of measurement reactivity and user experience dimensions using additional subscales extracted from validated questionnaires to measure stress, overload, usefulness, and satisfaction related to the use of the app.

Usability was measured using the SUS [48], as in study 1. The internal consistency of this scale in the study was α=.75.

Stress reactivity resulting from app use (subscale) was assessed through the Pressure/Tension subscale of the Intrinsic Motivation Inventory [52]. This subscale contains 5 items that measure the negative reactivity of participants while performing ESM with the tool. The internal consistency of this scale in the study was α=.56.

Overload resulting from app use (subscale) assessed the experienced negative affect and the degree of control and effort required during the completion of the ESM protocol. This was assessed using the Perceived Usability subscale from the User Engagement Scale [53], which comprises 8 items. The internal consistency of this scale in the study was α=.84.

To measure app usefulness (subscale), we used the Value subscale of the Intrinsic Motivation Inventory [52] to assess how participants found it useful to complete the ESM protocol by being more conscious of their own emotional and cognitive states. The internal consistency of this scale in the study was α=.94.

Satisfaction with the app was assessed using the Endurability subscale from the User Engagement Scale [53], referred to as the overall success of the interaction and users' willingness to recommend the app to others or engage with it in the future. The internal consistency of this scale in the study was α=.90.

App Compliance

Participants' level of commitment was also assessed in study 2. Therefore, latency and missing responses were recorded to perform compliance analyses, exactly as in study 1.

Analytic Plan

The analytic plan in study 1 was entirely reapplied in study 2. We conducted a descriptive analysis and performed Spearman correlation analyses because of the nonnormal distribution of the variables of interest. In addition, similar mixed model regression analyses were performed, as in study 1. Furthermore, we conducted an additional multivariate model in which we included methodological variables (day and time of the day), emotional symptomatology, usability, and variables related to the use and reactivity to the ESM system, as measured through the further included validated subscales (stress, overload, usefulness, and satisfaction). Therefore, separate analyses were conducted considering the item-based indices of measurement reactivity (ie, for the replication of the study 1 results) and further scale-based indices of measurement reactivity included in this study.

Ethics Approval

The study was approved by the Faculty Ethical Committee of Complutense University of Madrid (Protocol Code Ref. 2020/21-023) and complied with the Declaration of Helsinki's ethics standards.

Results

Sample Characteristics and General App Performance

As shown in Table 4, participants presented moderate depression and low anxiety levels, as measured with the CES-D 8 and GAD-7, respectively. Initially, we recruited 139 participants, from whom we found high compliance, as the level of abandonment was very low—2 (1.46%) participants in the sample (corresponding to 2 participants who completed <6 assessments). These 2 participants were excluded from the analysis because of the noncompletion of the questionnaires on usability and measurement reactivity at the end of the ESM protocol. Therefore, of the 139 participants, the final sample included in the analyses comprised 135 (97.1%) participants with a mean age of 20.51 (SD 2.31) years. We found a low mean number of missing responses per participant—3.60 (SD 3.32), ranging from 0 to 17 missing responses per participant. The mean levels of latency per participant found in this study were 23 to 79 (SD 8.1) minutes, ranging from a minimum of 6.15 minutes to a maximum of 49.64 minutes per participant.

Table 4. Sample characteristics and general app performance (study 2; N=135).

Characteristics	Values
Sex (female), n (%)	116 (85.9)
Age (years), mean (SD)	20.51 (2.31)
Anxiety (GAD-7[a]; 0-21), mean (SD)	8.38 (4.11)
Depression (CES-D[b] 8; 0-24), mean (SD)	9.09 (4.89)
App stress (item; 1-5), mean (SD)	2.96 (1.30)
App overload (item; 1-5), mean (SD)	2.93 (1.25)
App usefulness (item; 1-5), mean (SD)	3.44 (1.14)
Satisfaction (item; 1-10), mean (SD)	6.62 (1.92)
App stress (IMI[c] subscale; 1-7), mean (SD)	3.09 (1.09)
App overload (UES[d] subscale; 1-5), mean (SD)	1.86 (0.77)
App usefulness (IMI subscale; 1-7), mean (SD)	4.16 (1.37)
Satisfaction subscale (UES subscale; 1-5), mean (SD)	3.40 (0.95)
Usability (SUS[e]; 0-100), mean (SD)	82.17 (11.74)

[a]GAD-7: Generalized Anxiety Disorder-7.

[b]CES-D: Center for Epidemiologic Studies–Depression.

[c]IMI: Intrinsic Motivation Inventory.

[d]UES: User Engagement Scale.

[e]SUS: System Usability Scale.

In general, participants showed medium levels in variables related to the negative reactivity of the app (stress and overload), both measured using single items and their corresponding subscales. Participants showed medium to high scores in perceived usefulness when using the app and global satisfaction, as reflected by both the single items and the corresponding subscales. Overall, the results showed that participants did not report feeling stressed or overloaded because of the use of the system while participating in the study. Furthermore, participants found the app useful for its purpose of facilitating awareness of internal affective and cognitive states.

In addition, importantly, the usability of the app was rated as very high (mean 82.17, SD 11.74), as measured by SUS, where scores >80.30 are considered to be in 10% of the best-rated systems.

Correlation Analyses

The set of Spearman correlations is shown in Table S2 in Multimedia Appendix 1. As in study 1, we found a significant relationship between anxiety and depression levels (r=0.71; P<.001), indicating a certain degree of comorbidity among the symptoms. Second, we found a significant positive relationship between anxiety levels and reactive stress (measured by a single item; r=0.21; P=.01), indicating that participants who scored higher on anxiety levels felt more nervous or stressed because of ESM notifications. However, these results were not replicated in relation to the app stress subscale.

As in study 1, usability was significantly related to various variables concerning the reactivity with the app. We found negative significant relationships between usability and app stress (single item r=−0.21, P=.02; subscale r=−0.39, P<.001) and between usability and app overload (single item r=−0.40, P<.001; subscale r=−0.47, P<.001). We also found positive significant relationship between usability and app usefulness (single item r=0.26, P=.002; subscale r=0.43, P<.001) and app satisfaction (single item r=0.50, P<.001; subscale r=0.43, P<.001). Therefore, participants who found the app more usable showed lower levels of app stress and overload and higher levels of app usefulness and satisfaction with its use.

Variables referring to the negative measurement reactivity (stress and overload) were significantly and positively correlated (between single items: r=0.34, P<.001; between subscales: r=0.57, P<.001), indicating that those participants who felt more stressed because of notification requirements also felt higher overload because of completing each assessment. On the other hand, we found positive and significant correlations between general satisfaction with the app and perceived usefulness (between single items: r=0.45, P<.001; between subscales: r=0.72, P<.001). Overall, these results replicate and extend the previous findings in study 1 and are in line with previous research validating app-based tools of psychological assessment [54].

Multilevel Analyses

Latency

As in study 1, we conducted a series of transformations to control for the distribution of outliers. After that, we used linear mixed models, fitted by maximum likelihood estimation, to test time-, person-, and system-related predictors of the variability of latency when responding to experience sampling (results of empty univariate and multivariate models predicting latency can be found in Table 5). We used the same models as in study 1, and an additional multivariate model was tested, with variables related to measurement reactivity measured through the corresponding subscales. We also calculated the ICC for the empty model predicting latency responses, which showed a value of 22% for between-person variance.

Analyses testing the effects of design and time-related variables on latency showed a significant effect of the day number, showing the same effect in both the univariate (estimate 0.28 [SE 0.09]; P=.001) and multivariate models (estimate 0.29 [SE 0.09], P=.01; estimate 0.28 [SE 0.09], P=.001, respectively). Time of the day also showed significant effects on latency responses in the univariate model (estimate −0.42 [SE 0.21]; P=.04) and in both multivariate models (estimate −0.44 [SE 0.21], P=.04; estimate −0.43 [SE 0.21], P=.04, respectively). The results partially replicated those from study 1, showing that response latencies became longer as the study progressed; however, only in study 2, later notifications within the day were further related to lower latency rates and, therefore, to faster responses in those moments of the day.

As in study 1, analyses testing the effects of symptomatology levels on latency showed no significant effects of depression and anxiety on the latency response rates. Therefore, the levels of emotional symptomatology of the participants did not affect their latencies to respond to notifications.

Finally, analyses testing the effects of app-related variables on latency showed a significant effect of app overload, measured through the subscale (estimate 1.52 [SE 0.71]; P=.03) in the univariate model. This effect remained significant in the multivariate model when introduced with the remaining variables (estimate 2.20 [SE 1.02]; P=.03). Thus, higher experienced overload during the ESM protocol was associated with longer response latencies to notifications.

Table 5. Predictors of response latency (study 2).

Outcome variable: latency	Estimate (SE)	*t* value	*P* value
Empty model			
Intercept	23.76 (0.50)	47.9	<.001
Univariate models			
Day	0.28 (0.09)	3.22	.001
Time of day	−0.42 (0.21)	−2.04	.04
Anxiety	0.16 (0.12)	1.34	.18
Depression	−0.03 (0.10)	−0.33	.73
App stress item	0.36 (0.38)	0.95	.35
App overload item	0.54 (0.40)	1.37	.17
App usefulness item	−0.04 (0.43)	−0.08	.94
App satisfaction item	−0.313 (0.26)	−1.25	.21
App stress subscale	0.57 (0.47)	1.21	.23
App overload subscale	1.56 (0.66)	2.37	.02
App usefulness subscale	0.26 (0.37)	0.70	.48
App satisfaction subscale	0.05 (0.53)	0.10	.92
Usability	−0.01 (0.04)	−0.21	.83
Multivariate model 1 (item variables)			
Day	0.29 (0.09)	3.31	<.001
Time of day	−0.44 (0.21)	−2.1	.04
Anxiety	0.33 (0.17)	1.94	.05
Depression	−0.23 (0.14)	−1.59	.11
App stress item	0.05 (0.45)	0.11	.91
App overload item	0.34 (0.49)	0.70	.48
App usefulness item	0.20 (0.52)	0.39	.70
App satisfaction item	−0.39 (0.38)	−1.024	.31
Usability	0.04 (0.05)	0.71	.48
Multivariate model 2 (subscale variables)			
Day	0.28 (0.09)	3.25	.001
Time of day	−0.43 (0.21)	−2.07	.004
Anxiety	0.24 (0.17)	1.44	.15
Depression	−0.16 (0.14)	−1.16	.24
App stress subscale	−0.02 (0.59)	−0.03	.98
App overload subscale	2.39 (0.96)	2.50	.01
App usefulness subscale	0.10 (0.56)	0.18	.86
App satisfaction subscale	0.51 (0.83)	0.61	.54
Usability	0.06 (0.06)	0.99	.32

Missing

First, we calculated the empty model predicting missing responses and the ICC, which showed a value of 20% for between-person variance. Analyses testing the effects of design time–related variables on missing responses showed that time of the day was a significant predictor in both the univariate model (estimate 0.17 [SE 0.02]; *P*=.01) and the multivariate models 1 and 2 (estimate 0.16 [SE 0.06]; *P*=.01).

As for the analyses of the effects of emotional symptomatology on missing response rates, nonsignificant effects of depression and anxiety levels were found on missing responses.

Finally, analyses testing the relationship between app-related variables and missing response rates showed that item-based app stress had significant effects in the univariate model (estimate 0.21 [SE 0.07]; *P*=.004), which remained significant in multivariate model 1 (estimate 0.18 [SE 0.08]; *P*=.004). Item-based overload also showed a significant effect in the univariate model (estimate 0.16 [SE 0.08]; *P*=.04); however, this effect did not remain significant in multivariate model 1 (estimate 0.08 [SE 0.09]; *P*=.36). As for subscale-based measurements (model 2), app overload measured through the subscale also showed significant effects on missing rates when introduced as a single predictor (estimate 0.29 [SE 0.12]; *P*=.02), as well as on app satisfaction measured through a subscale (estimate –0.20 [SE 0.10]; *P*=.049). However, when all predictors were introduced in multivariate model 2, neither of these variables showed any significant effect (results of empty univariate and multivariate models predicting missing responses can be found in Table 6).

Table 6. Predictors of missing responses (study 2).

Outcome variable: missing	Estimate (SE)	z value	P value
Empty model			
Intercept	–2.43 (0.11)	–20.73	<.001
Univariate models			
Day	0.01 (0.02)	0.39	.70
Time of day	0.17 (0.02)	2.53	.01
Anxiety	0.03 (0.02)	1.39	.17
Depression	0.005 (0.02)	0.23	.82
App stress item	0.21 (0.07)	2.90	.004
App overload item	0.16 (0.08)	2.08	.04
App usefulness item	–0.06 (0.09)	–0.72	.47
App satisfaction item	–0.07 (0.05)	–1.45	.15
App stress subscale	0.10 (0.09)	1.14	.25
App overload subscale	0.29 (0.12)	2.37	.02
App usefulness subscale	–0.10 (0.07)	–1.42	.16
App satisfaction subscale	–0.20 (0.10)	–2.01	.04
Usability	–0.01 (0.01)	–1.48	.14
Multivariate model 1 (item variables)			
Day	0.01 (0.02)	0.41	.68
Time of day	0.16 (0.06)	2.54	.01
Anxiety	0.04 (0.03)	1.15	.25
Depression	–0.03 (0.03)	–1.01	.31
App stress item	0.18 (0.08)	2.11	.04
App overload item	0.08 (0.09)	0.91	.36
App usefulness item	–0.004 (0.10)	–0.04	.97
App satisfaction item	0.02 (0.07)	0.31	.76
Usability	–0.01 (0.01)	–0.50	.62
Multivariate model 2 (subscale variables)			
Day	0.01 (0.02)	0.39	.70
Time of day	0.16 (0.06)	2.55	.01
Anxiety	0.05 (0.03)	1.42	.16
Depression	–0.02 (0.03)	–0.84	.40
App stress subscale	–0.02 (0.11)	–0.20	.84
App overload subscale	0.21 (0.18)	1.17	.24
App usefulness subscale	–0.01 (0.10)	–0.14	.90
App satisfaction subscale	0.11 (0.16)	–0.71	.48
Usability	0.001 (0.01)	0.11	.91

Discussion (Study 2)

Study 2 was conducted to test the new ESM system using a longer ESM protocol than in study 1 and increase the methodological quality of app reactivity measurement from study 1, adding further assessments of these characteristics through validated subscales of app stress, overload, usefulness, and satisfaction.

Study 2 replicated the high usability scores of the new ESM system, with high levels over the 90th percentile of the SUS scale. Higher usability was related to several indicators such as lower user stress and overload and higher usefulness and

satisfaction, replicating and extending the findings from study 1 on the acceptability of the novel procedure.

Furthermore, compliance with the protocol was high, as in study 1, showing that extending the protocol regime did not affect general compliance. After performing mixed model analyses, we replicated the findings from study 1 on the time-related predictors of compliance with the procedure. We found a significant effect of day number on response latency, indicating that latencies became longer as the study progressed. These results are in line with the findings of study 1. In addition, a negative relationship between the time of the day and latency was found, indicating that in earlier notifications, latencies tended to be slower. This result was not found in study 1 and may indicate that extending the duration of the ESM protocol may permit the identification of performance patterns that may remain undetected for shorter durations (ie, a 5-day duration in study 1). Furthermore, we found that time of the day was also a significant predictor of missing responses, indicating more missing responses in later notifications of the day (interval between 8 PM and 9 PM). These results are in line with previous research showing lower compliance in terms of missing response rates in later daily notifications [28].

In addition, as in study 1, depression and anxiety levels were not found to significantly predict variability with ESM compliance, which is important for the future implementation of these measurements in clinical samples. Importantly, study 2 was completed from April to May 2021, when the COVID-19 pandemic was a persistent source of threat, and this was evidenced by participants' symptom levels, which were higher than those for participants in study 1, which was completed in 2019. Thus, although the samples were comparable in terms of sociodemographic characteristics, depression and anxiety reached higher overall subclinical levels in the sample of study 2. The fact that symptom levels in study 2 did not affect levels of compliance is clearly indicative of the feasibility of the ESM system for further implementation in clinical settings in the future.

In terms of measurement reactivity, we found significant effects of the measurement reactivity variables in predicting compliance in study 2. Higher app overload was related to more response latencies, whereas app stress significantly predicted more missing responses. These results are in line with previous research suggesting that a higher negative measurement reactivity is related to lower compliance rates [23]. Interestingly, these effects were not found in study 1, suggesting that the influence of these variables on the compliant performance of the new ESM system may only emerge when using more extensive protocols of assessment (ie, 10 days vs 5 days).

Discussion

Principal Findings

The aim of this study was to investigate the usability and feasibility of using a new ESM system, which integrates self-report questionnaires and an app-based cognitive task that allows the assessment of ecological indicators of cognitive-affective mechanisms implicated in emotion regulation and emotional symptoms, through 2 differently extensive ESM designs (studies 1 and 2). Study 1 required participants to respond to 3 assessments per day for 5 days. Study 2 integrated the ESM design into a protocol of assessments of 3 times a day for 10 days.

Across studies, we found similar results in terms of the average levels of usability of the novel ESM system, indicating its ease of use and learnability. This is particularly relevant because of the development of a new system that integrates an app-based cognitive task into the ESM procedure. Previous platforms have been developed to repeatedly assess affect through ESM, provide a visual environment [55-57], and highlight the importance of usability testing. In the case of the new ESM system used in our studies, the results on usability showed a very high level of learnability (ie, over the 90th percentile of the SUS), which reflects the ease of use and acceptability by the participants. In fact, we found that usability rates were significantly related to lower negative measurement reactivity (stress and overload) and more perceived usefulness and satisfaction with the app in both studies. Therefore, the use of this ESM system does not require an advanced technical background, and its implementation would be appropriate in the general population. The mean values of reactivity to the assessment measure were also similar across both studies, such that the levels of app overload were moderate (burdensome because of the length of assessments) and app usefulness was high (utility to increase users' conscientiousness of their internal psychological states). Furthermore, person-related variables of depression and anxiety did not affect compliance with the ESM protocol. This also informs on the feasibility of further expanding and implementing the new ESM system in clinical settings.

Moreover, we found different results between studies, such that only in study 2 (ie, with a more extensive protocol of 10 days), individual differences in app overload and stress accounted for lower compliance rates. However, it must be noted that the degree of between-person variability in compliance rates was relatively low (as indicated by the ICC derived from empty models), suggesting that such effects might turn out anecdotical.

Limitations

Despite the implications of these studies, some limitations must be considered. First, study 2 was conducted during the COVID-19 pandemic, which makes it difficult to generalize the results in terms of compliance with other regular contexts of evaluation, given the exceptional circumstances in which the participants found themselves at that time. Despite this, measures of reactivity to the mobile app and its usability remained very similar between both studies, which indicates that they may depend more on other design- or sample-related characteristics rather than on contextual conditions. This implies the feasibility of applying similar designs with the new ESM system in different contexts in a reliable manner, including the use of the novel app for the study of cognitive and emotional dynamics within individuals across multiple contexts across time. Furthermore, in terms of between-person differences, the relatively high scores in depression and anxiety levels in the sample in study 2 (subclinical levels) minimally affected compliance with the app, which is indicative of its feasibility

for further implementation with other types of at-risk and clinical populations.

Second, studies 1 and 2 differ in the way the ESM system was implemented, as it was used through a mobile app in the former one, whereas in the latter study, an integrated app system that could be completed both on phones and computers was used. Importantly, despite this difference, participants showed similar scores on usability and measurement reactivity with the system, pointing to the adequacy of using either format of the new ESM system, depending on specific user requirements.

Third, measurement reactivity variables were only measured through item-based assessments in study 1. This was solved in study 2, in which we added further validated subscales to measure the reactivity variables. Although these subscales resembled the constructs gathered by the single items, the results were not fully replicated in multivariable models 1 (item-based) and 2 (scale-based) of study 2. As these scales showed high reliability in general and added higher methodological quality, it is recommended to use them to replicate results in future studies.

Future research should also focus on incorporating other cognitive evaluation tasks into ESM systems, not just self-report assessments, combining both implicit and explicit assessments of emotional, behavioral, and cognitive processes. This promises to bring great wealth to the understanding of the psychological dynamic underlying the ecological mechanisms of people's emotional dysfunctions (ie, depression and anxiety) in daily life.

Conclusions

This study supports the validity and feasibility of the presented new ESM system. Furthermore, our findings indicate that more systematic investigations into the design characteristics influencing data quality and quantity in ESM studies are needed. The variability of compliance rates, in terms of latency and missing responses, depended on variables related to the design of the ESM procedures and measurement reactivity variables in our studies. Lower compliance was found across both studies as the days in each study progressed, and measurement reactivity variables were found to be related to lower compliance rates (higher latency and missing responses) in more extensive ESM protocols (study 2). Importantly, participants' levels of depressive and anxiety symptomatology did not affect compliance in our study, indicating the feasibility of using this type of new system in ESM designs for multiple populations. This will permit the evaluation of ecological, cognitive, and emotional dynamics in individuals' daily life and a better understanding of the underlying mechanisms ultimately influencing mental health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Experience sampling method assessments and correlation analyses tables.
[DOCX File , 169 KB - formative_v6i3e32537_app1.docx]

References

1. Fried EI, Cramer AO. Moving forward: challenges and directions for psychopathological network theory and methodology. Perspect Psychol Sci 2017 Nov 05;12(6):999-1020. [doi: 10.1177/1745691617705892] [Medline: 28873325]
2. Beck AT. Cognitive therapy and the emotional disorders. Am J Physio Ther 1976 Oct;31(4):633-634. [doi: 10.1176/appi.psychotherapy.1977.31.4.633]
3. American Psychiatric Association. Diagnostic And Statistical Manual Of Mental Disorders, Fifth Edition. Washington, DC: American Psychiatric Association; 2013.
4. Fisher AJ. Toward a dynamic model of psychological assessment: implications for personalized care. J Consult Clin Psychol 2015 Aug;83(4):825-836. [doi: 10.1037/ccp0000026] [Medline: 26009781]
5. Barrett LF, Barrett DJ. An introduction to computerized experience sampling in psychology. Soc Sci Comput Rev 2016 Aug 18;19(2):175-185 [FREE Full text] [doi: 10.1177/089443930101900204]

XSL•FO
RenderX

6. Cutler SE, Larsen RJ, Bunce SC. Repressive coping style and the experience and recall of emotion: a naturalistic study of daily affect. J Pers 1996 Jun;64(2):379-405. [doi: 10.1111/j.1467-6494.1996.tb00515.x] [Medline: 8656322]

7. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008 Apr;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]

8. Bylsma LM, Taylor-Clift A, Rottenberg J. Emotional reactivity to daily events in major and minor depression. J Abnorm Psychol 2011 Feb;120(1):155-167. [doi: 10.1037/a0021662] [Medline: 21319928]

9. Brans K, Koval P, Verduyn P, Lim YL, Kuppens P. The regulation of negative and positive affect in daily life. Emotion 2013 Oct;13(5):926-939. [doi: 10.1037/a0032400] [Medline: 23731436]

10. Ford BQ, Gross JJ, Gruber J. Broadening our field of view: the role of emotion polyregulation. Emot Rev 2019 Jun 23;11(3):197-208. [doi: 10.1177/1754073919850314]

11. Haines SJ, Gleeson J, Kuppens P, Hollenstein T, Ciarrochi J, Labuschagne I, et al. The wisdom to know the difference. Psychol Sci 2016 Dec 15;27(12):1651-1659. [doi: 10.1177/0956797616669086] [Medline: 27738099]

12. Armstrong T, Olatunji BO. Eye tracking of attention in the affective disorders: a meta-analytic review and synthesis. Clin Psychol Rev 2012 Dec;32(8):704-723 [FREE Full text] [doi: 10.1016/j.cpr.2012.09.004] [Medline: 23059623]

13. Sanchez A, Vazquez C, Marker C, LeMoult J, Joormann J. Attentional disengagement predicts stress recovery in depression: an eye-tracking study. J Abnorm Psychol 2013 May;122(2):303-313. [doi: 10.1037/a0031529] [Medline: 23421524]

14. Duque A, Vázquez C. Double attention bias for positive and negative emotional faces in clinical depression: evidence from an eye-tracking study. J Behav Ther Exp Psychiatry 2015 Mar;46:107-114. [doi: 10.1016/j.jbtep.2014.09.005] [Medline: 25305417]

15. Everaert J, Grahek I, Duyck W, Buelens J, Van den Bergh N, Koster EH. Mapping the interplay among cognitive biases, emotion regulation, and depressive symptoms. Cogn Emot 2017 Jun 16;31(4):726-735. [doi: 10.1080/02699931.2016.1144561] [Medline: 26878897]

16. Blanco I, Vazquez C. Integrative well-being leads our attentional system: an eye-tracking study. J Happiness Stud 2021 Apr 06;22(2):787-801. [doi: 10.1007/s10902-020-00251-7]

17. Wenzlaff RM, Bates DE. Unmasking a cognitive vulnerability to depression: how lapses in mental control reveal depressive thinking. J Personal Soc Psychol 1998;75(6):1559-1571. [doi: 10.1037/0022-3514.75.6.1559]

18. Sanchez-Lopez A, Everaert J, Van Put J, De Raedt R, Koster E. Eye-gaze contingent attention training (ECAT): examining the causal role of attention regulation in reappraisal and rumination. Biol Psychol 2019 Mar;142:116-125 [FREE Full text] [doi: 10.1016/j.biopsycho.2019.01.017] [Medline: 30735680]

19. Sanchez-Lopez A, De Raedt R, van Put J, Koster EH. A novel process-based approach to improve resilience: effects of computerized mouse-based (gaze)contingent attention training (MCAT) on reappraisal and rumination. Behav Res Ther 2019 Jul;118:110-120. [doi: 10.1016/j.brat.2019.04.005] [Medline: 31048096]

20. Blanco I, Boemo T, Sanchez-Lopez A. An online assessment to evaluate the role of cognitive biases and emotion regulation strategies for mental health during the COVID-19 lockdown of 2020: structural equation modeling study. JMIR Ment Health 2021 Nov 02;8(11):e30961 [FREE Full text] [doi: 10.2196/30961] [Medline: 34517337]

21. Myin-Germeys I, Oorschot M, Collip D, Lataster J, Delespaul P, van Os J. Experience sampling research in psychopathology: opening the black box of daily life. Psychol Med 2009 Feb 12;39(9):1533-1547. [doi: 10.1017/s0033291708004947]

22. Palmier-Claus J, Myin-Germeys I, Barkus E, Bentley L, Udachina A, Delespaul PA, et al. Experience sampling research in individuals with mental illness: reflections and guidance. Acta Psychiatr Scand 2011 Jan;123(1):12-20. [doi: 10.1111/j.1600-0447.2010.01596.x] [Medline: 20712828]

23. Eisele G, Vachon H, Lafit G, Kuppens P, Houben M, Myin-Germeys I, et al. The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. Assessment 2020 Sep 10:1073191120957102. [doi: 10.1177/1073191120957102] [Medline: 32909448]

24. Sun J, Rhemtulla M, Vazire S. Eavesdropping on missing data: what are university students doing when they miss experience sampling reports? Pers Soc Psychol Bull 2020 Nov;47(11):1535-1549. [doi: 10.1177/0146167220964639] [Medline: 33342369]

25. Broderick JE, Schwartz JE, Shiffman S, Hufford MR, Stone AA. Signaling does not adequately improve diary compliance. Ann Behav Med 2003 Oct;26(2):139-148. [doi: 10.1207/S15324796ABM2602_06] [Medline: 14534031]

26. Litcher-Kelly L, Kellerman Q, Hanauer SB, Stone AA. Feasibility and utility of an electronic diary to assess self-report symptoms in patients with inflammatory bowel disease. Ann Behav Med 2007 Apr;33(2):207-212. [doi: 10.1007/BF02879902] [Medline: 17447873]

27. Ono M, Schneider S, Junghaenel DU, Stone AA. What affects the completion of ecological momentary assessments in chronic pain research? An individual patient data meta-analysis. J Med Internet Res 2019 Feb 05;21(2):e11398 [FREE Full text] [doi: 10.2196/11398] [Medline: 30720437]

28. Rintala A, Wampers M, Myin-Germeys I, Viechtbauer W. Response compliance and predictors thereof in studies using the experience sampling method. Psychol Assess 2019 Feb;31(2):226-235. [doi: 10.1037/pas0000662] [Medline: 30394762]

29. Silvia PJ, Kwapil TR, Eddington KM, Brown LH. Missed beeps and missing data: dispositional and situational predictors of nonresponse in experience sampling research. Soc Sci Comput Rev 2013 Mar 13;31(4):471-481. [doi: 10.1177/0894439313479902]

30. Sokolovsky AW, Mermelstein RJ, Hedeker D. Factors predicting compliance to ecological momentary assessment among adolescent smokers. Nicotine Tob Res 2014 Mar;16(3):351-358 [FREE Full text] [doi: 10.1093/ntr/ntt154] [Medline: 24097816]

31. Johnson EI, Grondin O, Barrault M, Faytout M, Helbig S, Husky M, et al. Computerized ambulatory monitoring in psychiatry: a multi-site collaborative study of acceptability, compliance, and reactivity. Int J Methods Psychiatr Res 2009 Mar;18(1):48-57 [FREE Full text] [doi: 10.1002/mpr.276] [Medline: 19195050]

32. Jones A, Remmerswaal D, Verveer I, Robinson E, Franken IH, Wen CK, et al. Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. Addiction 2019 Apr 21;114(4):609-619 [FREE Full text] [doi: 10.1111/add.14503] [Medline: 30461120]

33. Vachon H, Viechtbauer W, Rintala A, Myin-Germeys I. Compliance and retention with the experience sampling method over the continuum of severe mental disorders: meta-analysis and recommendations. J Med Internet Res 2019 Dec 06;21(12):e14475 [FREE Full text] [doi: 10.2196/14475] [Medline: 31808748]

34. Messiah A, Grondin O, Encrenaz G. Factors associated with missing data in an experience sampling investigation of substance use determinants. Drug Alcohol Depend 2011 Apr 01;114(2-3):153-158. [doi: 10.1016/j.drugalcdep.2010.09.016] [Medline: 21075563]

35. Rintala A, Wampers M, Myin-Germeys I, Viechtbauer W. Momentary predictors of compliance in studies using the experience sampling method. Psychiatry Res 2020 Feb 29;286:112896. [doi: 10.1016/j.psychres.2020.112896] [Medline: 32146247]

36. Conner TS, Reid KA. Effects of intensive mobile happiness reporting in daily life. Soc Psychol Personal Sci 2011 Aug 29;3(3):315-323. [doi: 10.1177/1948550611419677]

37. Hufford MR, Shields AL, Shiffman S, Paty J, Balabanis M. Reactivity to ecological momentary assessment: an example using undergraduate problem drinkers. Psychol Addict Behav 2002 Sep;16(3):205-211. [doi: 10.1037/0893-164X.16.3.205]

38. Arslan R, Reitz A, Driebe J, Gerlach T, Penke L. Routinely randomize potential sources of measurement reactivity to estimate and adjust for biases in subjective reports. PsyArXiv. 2020. URL: https://psyarxiv.com/va8bx/ [accessed 2022-03-10]

39. Chiu TM, Eysenbach G. Stages of use: consideration, initiation, utilization, and outcomes of an internet-mediated intervention. BMC Med Inform Decis Mak 2010;10:73 [FREE Full text] [doi: 10.1186/1472-6947-10-73] [Medline: 21092275]

40. Lenert L, Muñoz RF, Stoddard J, Delucchi K, Bansod A, Skoczen S, et al. Design and pilot evaluation of an internet smoking cessation program. J Am Med Inform Assoc 2003 Jan 01;10(1):16-20 [FREE Full text] [doi: 10.1197/jamia.m1128] [Medline: 12509354]

41. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. Appl Psychol Measur 2016 Jul 26;1(3):385-401. [doi: 10.1177/014662167700100306]

42. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

43. Karim J, Weisz R, Bibi Z, ur Rehman S. Validation of the eight-item center for epidemiologic studies depression scale (CES-D) among older adults. Curr Psychol 2014 Oct 22;34(4):681-692. [doi: 10.1007/s12144-014-9281-y]

44. Van de Velde S, Levecque K, Bracke P. Measurement equivalence of the CES-D 8 in the general population in Belgium: a gender perspective. Arch Public Health 2009 Apr 20;67(1):15. [doi: 10.1186/0778-7367-67-1-15]

45. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]

46. Ruiz MA, Zamorano E, García-Campayo J, Pardo A, Freire O, Rejas J. Validity of the GAD-7 scale as an outcome measure of disability in patients with generalized anxiety disorders in primary care. J Affect Disord 2011 Feb;128(3):277-286. [doi: 10.1016/j.jad.2010.07.010] [Medline: 20692043]

47. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]

48. Brooke J. SUS-A quick and dirty usability scale. In: Jordan P, Lyall I, Weerdmeester M, editors. Usability Evaluation In Industry. Boca Raton, Florida: CRC Press; 1996:189-194.

49. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum-Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]

50. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using. J Stat Soft 2015;67(1). [doi: 10.18637/jss.v067.i01]

51. Lo S, Andrews S. To transform or not to transform: using generalized linear mixed models to analyse reaction time data. Front Psychol 2015 Aug;6:1171 [FREE Full text] [doi: 10.3389/fpsyg.2015.01171] [Medline: 26300841]

52. McAuley E, Duncan T, Tammen VV. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: a confirmatory factor analysis. Res Q Exerc Sport 1989 Mar;60(1):48-58. [doi: 10.1080/02701367.1989.10607413] [Medline: 2489825]

53. O'Brien HL, Toms EG. The development and evaluation of a survey to measure user engagement. J Am Soc Inf Sci 2009 Oct 19;61(1):50-69. [doi: 10.1002/asi.21229]

54. Daniëls NE, Bartels S, Verhagen S, Van Knippenberg R, De Vugt M, Delespaul P. Digital assessment of working memory and processing speed in everyday life: feasibility, validation, and lessons-learned. Internet Interv 2020 Mar;19:100300 [FREE Full text] [doi: 10.1016/j.invent.2019.100300] [Medline: 31970080]

55. Bailon C, Damas M, Pomares H, Sanabria D, Perakakis P, Goicoechea C, et al. Smartphone-based platform for affect monitoring through flexibly managed experience sampling methods. Sensors (Basel) 2019 Aug 05;19(15):3430 [FREE Full text] [doi: 10.3390/s19153430] [Medline: 31387292]

56. Rough D, Quigley A. Jeeves - A visual programming environment for mobile experience sampling. In: Proceedings of the IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC). 2015 Presented at: IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC); Oct. 18-22, 2015; Atlanta, GA, USA. [doi: 10.1109/vlhcc.2015.7357206]

57. Rough D, Quigley A. Challenges of traditional usability evaluation in end-user development. In: Proceedings of the International Symposium on End User Development. 2019 Presented at: International Symposium on End User Development; July 10-12, 2019; Hatfield, United Kingdom p. 1-17. [doi: 10.1007/978-3-030-24781-2_1]

Abbreviations

CES-D: Center for Epidemiologic Studies–Depression
ESM: experience sampling method
GAD-7: Generalized Anxiety Disorder-7
ICC: interclass correlation coefficient
SST: Scrambled Sentence Task
SUS: System Usability Scale

[XSL•FO]

RenderX

<u>Original Paper</u>

Using Named Entity Recognition to Identify Substances Used in the Self-medication of Opioid Withdrawal: Natural Language Processing Study of Reddit Data

Alexander Preiss[1], MS; Peter Baumgartner[1,2], MS; Mark J Edlund[3], MD, PhD; Georgiy V Bobashev[1], PhD

[1]Center for Data Science, RTI International, Durham, NC, United States

[2]ExplosionAI GmbH, Berlin, Germany

[3]Community Health Research Division, RTI International, Durham, NC, United States

Corresponding Author:
Alexander Preiss, MS
Center for Data Science
RTI International
3040 E Cornwallis Rd
Durham, NC, 27709
United States
Phone: 1 9195418771
Email: preisssa@gmail.com

Abstract

Background: The cessation of opioid use can cause withdrawal symptoms. People often continue opioid misuse to avoid these symptoms. Many people who use opioids self-treat withdrawal symptoms with a range of substances. Little is known about the substances that people use or their effects.

Objective: The aim of this study is to validate a methodology for identifying the substances used to treat symptoms of opioid withdrawal by a community of people who use opioids on the social media site Reddit.

Methods: We developed a named entity recognition model to extract substances and effects from nearly 4 million comments from the *r/opiates* and *r/OpiatesRecovery* subreddits. To identify effects that are symptoms of opioid withdrawal and substances that are potential remedies for these symptoms, we deduplicated substances and effects by using clustering and manual review, then built a network of substance and effect co-occurrence. For each of the 16 effects identified as symptoms of opioid withdrawal in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, we identified the 10 most strongly associated substances. We classified these pairs as follows: substance is a Food and Drug Administration–approved or commonly used treatment for the symptom, substance is not often used to treat the symptom but could be potentially useful given its pharmacological profile, substance is a home or natural remedy for the symptom, substance can cause the symptom, or other or unclear. We developed the Withdrawal Remedy Explorer application to facilitate the further exploration of the data.

Results: Our named entity recognition model achieved F_1 scores of 92.1 (substances) and 81.7 (effects) on hold-out data. We identified 458 unique substances and 235 unique effects. Of the 130 potential remedies strongly associated with withdrawal symptoms, 54 (41.5%) were Food and Drug Administration–approved or commonly used treatments for the symptom, 17 (13.1%) were not often used to treat the symptom but could be potentially useful given their pharmacological profile, 13 (10%) were natural or home remedies, 7 (5.4%) were causes of the symptom, and 39 (30%) were other or unclear. We identified both potentially promising remedies (eg, gabapentin for body aches) and potentially common but harmful remedies (eg, antihistamines for restless leg syndrome).

Conclusions: Many of the withdrawal remedies discussed by Reddit users are either clinically proven or potentially useful. These results suggest that this methodology is a valid way to study the self-treatment behavior of a web-based community of people who use opioids. Our Withdrawal Remedy Explorer application provides a platform for using these data for pharmacovigilance, the identification of new treatments, and the better understanding of the needs of people undergoing opioid withdrawal. Furthermore, this approach could be applied to many other disease states for which people self-manage their symptoms and discuss their experiences on the web.

XSL•FO
RenderX

KEYWORDS

substance abuse; opioid epidemic; opioid use disorder; self-medication; social media; Reddit; natural language processing; machine learning; network analysis; opioid; drug withdrawal; withdrawal; opioid withdrawal; mobile phone

Introduction

Background

Withdrawal symptoms are major contributing factors of continued opioid misuse and relapse among those who attempt to quit [1-5]. Opioid-related withdrawal can be severe and may last for a week or longer. The symptoms often include body ache, diarrhea, nausea and vomiting, profuse sweating, insomnia, and loss of appetite [6,7]. Many patients relapse to using opioids to alleviate the symptoms. The medical treatment of withdrawal is a critical target in opioid treatment—the National Institute on Drug Abuse identified finding new treatments for opioid use disorders as its highest priority [8].

Clinicians use a variety of treatments for opioid withdrawal symptoms. Currently, opioid agonists such as methadone and buprenorphine are the most common treatments for opioid use disorder, and these are often tapered, during which time withdrawal symptoms may occur. These symptoms are often managed with standard treatments such as loperamide for diarrhea, ibuprofen for body aches, and odansetron for nausea. In 2018, the US Food and Drug Administration (FDA) approved Lucemyra (lofexidine) as the first nonopioid medication focused specifically on treating withdrawal symptoms [9,10]. At the same time, some physicians prescribe medications off-label (eg, baclofen and clonidine) to treat withdrawal. Often, these medications are used for relief of specific symptoms, such as nausea, diarrhea, or body aches. For some of these off-label treatments, the evidence base is good, but it is not as strong for others [11-15].

Although only a fraction of opioid users seek professional help to mitigate withdrawal symptoms, many seek advice from other opioid users through web-based forums and blogs [16-18]. Opioid users are actively experimenting with alternative treatments to alleviate their withdrawal symptoms. These *remedies* include use of over-the-counter medications such as loperamide for diarrhea; more experimental medication trials such as supplements (eg, vitamins and herbs); and other methods such as meditation, yoga, and acupuncture. Some of these alternative treatments are controversial (eg, the use of the opioid-containing food supplement kratom). The self-treatment practices of opioid users are poorly understood.

Social media offers unique insights into millions of web-based conversations about withdrawal remedies and can be analyzed using machine learning techniques. The broad involvement of the middle-class population in the opioid epidemic combined with the popularity of social media and the availability of smartphone devices has made web-based discussion of opioid use common. Several outlets provide searchable and analyzable information suitable for research: web-based forums such as Reddit and Bluelight, smaller personal blogs, support groups, and treatment centers, as well as Twitter. The amount of information about drug use and drug recovery contained in discussion forums is unparalleled; nowhere else is it possible

to obtain such rich information about drug use and drug recovery practices. Recent studies have analyzed forum data related to opioid recovery [19], buprenorphine [20], marijuana [21], social networking [22], and emerging trends in drug use [23]. Others have shown that web-based discussion of opioids correlates with key surveillance metrics, such as synthetic opioid death rates, and could be used as a leading indicator [24]. Although studies have begun to use these sources, such information remains underused. There have been no assessments of substances used for relieving withdrawal symptoms.

Objectives

The purpose of this study is 2-fold: (1) to validate a methodology that uses social media (Reddit posts) to investigate these self-treatment practices and (2) to better understand these practices, such as what is being used and what the consequences are of such self-help. We note that a validated methodology that uses Reddit posts to understand issues such as self-medication could have utility for a number of physical and behavioral disorders.

In our study, we focus on the following two Reddit discussion forums: *r/opiates* ("discussion of all things related to the narcotics known as opiates, from harm-reduction to pharmacology") and *r/OpiatesRecovery* ("...a group of people dedicated to helping each other kick the habit") [16,18]. Both forums are dedicated to open dialogue about opioid use, often with the intent of helping current and past users recover. As a part of these discussions, users often share their experiences with formal treatments and alternative treatments to mitigate the effects of withdrawal.

Our primary objective is to validate a methodology for identifying substances used to treat withdrawal symptoms from the discussions on these forums. There are no validation standards to relate discussions to the actual prevalence, incidence, and more detailed representative epidemiology of use, in large part because the forums are anonymous. At the same time, discussion forums can provide insights on the general knowledge among people who use opioids regarding the pharmacology of drugs that they are prescribed, that they buy over the counter, or that they obtain illicitly. We implicitly assume that, because of the large volume of discussions on social media, the strongest signals of remedies associated with prominent withdrawal symptoms would be clinically useful. We further expect that *common knowledge* would be more prevalent than *urban myths*. In addition to validating the methodology, we intend to demonstrate its utility for identifying clearly harmful approaches and discovering new, potentially useful remedies for withdrawal symptoms. If successful, such an approach may be useful to investigate other related questions about substance use, temporal trends, polydrug use, and other disorders where people self-manage their symptoms and discuss their experiences on the web.

Recent advances in natural language processing (NLP) have enabled researchers to identify and extract information from

large amounts of text, including useful information about potential remedies. NLP has been applied to an array of problems across the health care and public health fields [25], including many aspects of the opioid epidemic [19]. Using named entity recognition (NER), we train a model on a subsample of annotated discussions from both subreddits and then apply the model to extract entities from the rest of the data. We train the NER model to identify *substances* and *effects,* where effects are *the result of using or not using a substance.* In this analysis, we focus on the effects that are symptoms of opioid withdrawal and substances that are potential remedies for those withdrawal symptoms. Thus, effects are categorized as (1) symptoms of opioid withdrawal in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), (eg, body aches); (2) effects of opioid use (eg, euphoria); (3) a medical symptom not falling into categories 1 or 2; or (4) other or unclear. Finally, we conducted a validation of the strongest signals in the data. We also developed a withdrawal remedy database and an exploration-discovery tool to assist clinicians and the pharmaceutical industry to better serve the needs of people who use opioids.

This paper is structured as follows. First, we describe the data, entity definitions, and approaches for data collection. We then describe the process for iteratively training an NER model, using it to assemble an extensive database of opioid-related substances and effects and cleaning the resulting data. Next, we describe a network analysis approach to structure the database as a network of substance and effect co-occurrence and the development of a web application tool for exploring the database. Finally, we describe the validation process, where we systematically assess some of the strongest signals in the data. We conclude with a discussion of the value of our approach, database, and tool and consider the next steps in this research.

Methods

Data

Data Acquisition and Preparation

Reddit is a public social media site comprising communities called subreddits, which organize content based on interest. A submission to Reddit is called a *post*. A post can be a link to a website outside of Reddit or a piece of text for discussion. In the latter case, this is called a *selfpost*, and the text is called *selftext*. On a post, other Reddit users can provide threaded comments. Posts and comments are archived via the pushshift service [26], and a publicly available copy of the pushshift Reddit data set is also available via the Google BigQuery platform. Using BigQuery, we downloaded all available posts and comments from the *r/opiates* (when referring to the subreddits, we adopt the nomenclature of prefixing them with *r/*) and *r/OpiatesRecovery* subreddits (the corpus), which at the time of acquisition (August 2019) extended until April 2019. In October 2020, we downloaded additional data covering May-December 2019 directly from pushshift.io using the *PushshiftRedditDistiller* package [27].

A submission on Reddit contains three possible sources of text for analysis: the title of the post, the *selftext* of the post (if it is a *selfpost*), and the threaded comments for a post. As our goal was to detect mentions of our entities within longer-form text and post titles and *selftext* often contain short phrases or incomplete sentences, we focused only on the comments that appear as discussions on a post and did not use text from post titles or *selftext* for analysis.

r/opiates and r/OpiatesRecovery Subreddits

We focused on content from two communities related to opioid use*:* the *r/opiates* and *r/OpiatesRecovery* subreddits. The *r/opiates* subreddit was created on June 24, 2009, and *r/OpiatesRecovery* was created on February 16, 2012. As of June 25, 2021, the *r/opiates* subreddit had 124,696 members, and *r/OpiatesRecovery* had 31,522 members [16,18].

Some basic statistics about the comments within each subreddit are presented in Table 1. Users have the ability to delete specific comments or delete their accounts and all comments, so both the count and nondeleted count are presented.

Table 1. Summary of comments from *r/opiates* and *r/OpiatesRecovery*.

Item	Subreddit		Total
	r/opiates	*r/OpiatesRecovery*	
First comment, date; time	April 8, 2010; 4:10 AM	February 16, 2012; 5:19 AM	N/A[a]
Last comment, date; time	December 31, 2019; 11:57 PM	December 31, 2019; 11:58 PM	N/A
Count	3,650,602	341,598	3,992,200
Count nondeleted	3,446,046	326,729	3,772,775

[a]N/A: not applicable.

NER Model

We framed our NLP task as an NER problem to identify possible remedies and their effects. We aimed to identify two possible types of entities: substances and effects.

- Substance: a drug, remedy, supplement, or other consumable item used to treat an effect (eg, acetaminophen to treat body aches) or induce a desired effect (eg, heroin to induce euphoria). Although we found mentions of

meditation, yoga, and other nonmedicinal remedies, in this report we only focus on consumable substances.

- Effect: a negative or positive effect mentioned as a result of consuming a substance (eg, constipation caused by opioid use or constipation relieved by polyethylene glycol use), a result of not consuming a substance (eg, diarrhea caused by opioid withdrawal), or a rationale for consuming a substance (eg, diarrhea prompting loperamide use).

We used an iterative data-labeling and model-training process to generate 6507 labeled comments that formed our training data set. For details, see Multimedia Appendix 1 [28-36]. We trained our final NER model with the default spaCy (version 2.3.0; Explosion AI) [37] settings for NER models: 10 epochs with a dropout of 0.2 [38] and compounding batch. We trained the model on 80% of the data and evaluated its performance on a randomly selected hold-out set of 20%. Precision, recall, and F1 scores were calculated for exact matches of entities.

Our complete data-processing, modeling, and analysis pipeline is shown in Figure 1 and described in detail in Multimedia Appendix 1.

Figure 1. Withdrawal remedy analysis pipeline. NER: named entity recognition.
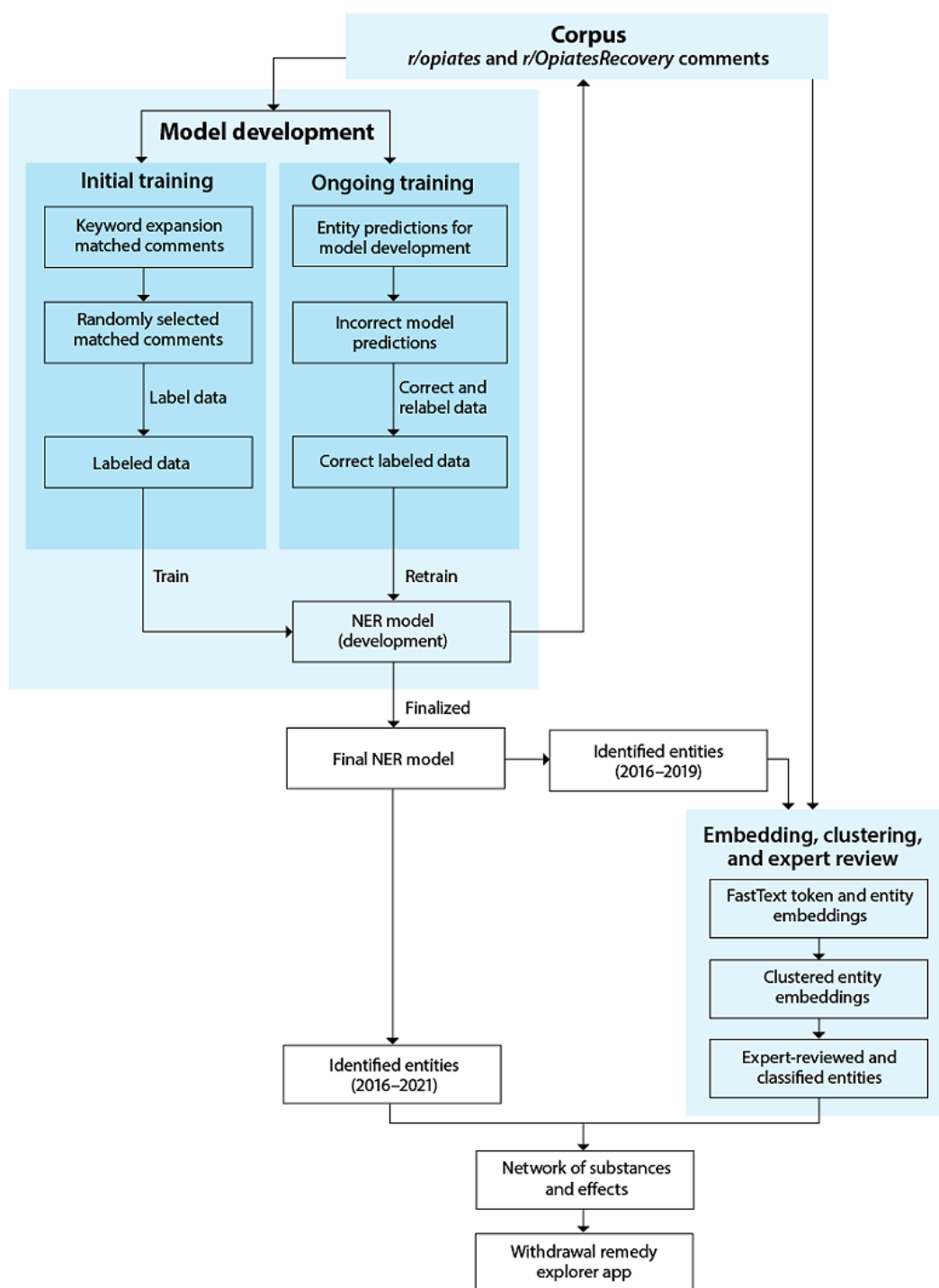
Deduplication of Entities

Many of the entities identified by the model were either misspellings or semantic duplicates. Misspellings were common in the data (eg, various misspellings of a drug name). Semantic duplicates resulted when the data included semantically related words, such as synonyms, hypernyms, and hyponyms. Some of the most common semantic duplicate types in our entities included the following: (1) slang terms (eg, *fent* for *fentanyl*), (2) generic and brand name drugs (eg, *loperamide* and *Imodium*), and (3) specific and general descriptions of an effect (eg, *insomnia* and *trouble sleeping*).

To deduplicate the entities, we generated fastText word embeddings for all tokens, where tokens identified as entities were combined [39]. We then clustered entities with similar embeddings. This approach is technically related (though with opposite goals) to embedding-based methods for vocabulary expansion, which have recently been applied to opioid-related discussions from Reddit [40] (see *Detailed Methods* in Multimedia Appendix 1). After clustering, all entities in a cluster were replaced with the most common entity in the cluster. For example, *cigarettes*, *ciggarettes* [*sic*], *ciggs*, and *cigs* were clustered. The most common entity in this cluster was *cigarettes*. All occurrences of the other entities in the cluster were replaced with *cigarettes*.

Expert Review and Classification of Entities

After clustering, we manually reviewed the deduplicated entities and conducted an additional round of deduplication and classification. Clustering was essential to reduce the number of entities and make manual review feasible. We generated separate lists of substances and the effects of these substances. The lists were reviewed by a psychiatrist (MJE) in 2 steps. First, we gave each entity a corrected name. This provided further deduplication of any slang terms or misspellings that remained after clustering. It also ensured that all entities were given a clinically accurate name. Often, these were controlled substances (eg, opioids, benzodiazepines, or prescription stimulants), illicit drugs (eg, hallucinogens or marijuana), or medications used to treat the symptoms of opioid withdrawals (eg, antiemetic or antidiarrheal medications). Second, we categorized the entities. Substances were categorized based on their pharmacological profile and use in clinical practice. Effects were categorized as (1) DSM-5 symptoms of opioid withdrawal (eg, body aches), (2) effects of opioid use (eg, euphoria), (3) a medical symptom not falling into categories 1 or 2 (eg, seizure), or (4) other or unclear (eg, sickness).

Network Analysis

We generated a bipartite network of substance and effect co-occurrence to assess which substances were associated with these effects. We considered entities to be nodes in the network. We drew edges between nodes when a substance and effect co-occurred in a sentence. We weighted the edges by the number of times the node pair co-occurred. However, using this edge weight alone to identify significant substance-effect pairs would favor pairs where the individual probabilities of occurrence of the substance and effect were high. Therefore, we also weighted the edges using positive pointwise mutual information (PPMI) [41]. For a pair of nodes, PPMI is high when their probability of co-occurrence is high relative to their individual probabilities of occurrence.

Application Development: Withdrawal Remedy Explorer

We built a web application called Withdrawal Remedy Explorer to provide a user-friendly way to explore the substance-effect network. Owing to its size, it was difficult to interpret the full network. Therefore, our application allows users to filter down to the ego network for a single substance or effect. Ego networks consist of a single node (ego) together with the nodes it is connected to (alters) [42] (in most cases, ego networks also include connections between the alters but, because our network is bipartite, there are no connections between entities of the same type). The application further allows users to filter according to edge count and PPMI.

Validation Review of Pairs

We generated and validated a list of top symptom and potential remedy pairs to assess the nature of the network. For each of the 16 effects identified as DSM-5 symptoms of opioid withdrawal, we identified the top 10 nonopioid substances most strongly associated (although not necessarily causally) with the symptom. We identified these substances using a weighted average of edge count and PPMI. First, we took the natural logarithm of the edge count to reduce skewness. We then used a min–max scaler to normalize the edge count and PPMI to a range of 0 to 1. Finally, we averaged the scaled edge count and PPMI. For each withdrawal symptom, we took the 10 substances with the highest value for this calculation, omitting opioids and any pairs with an edge count <5. This produced a list of 130 strongly associated symptom and potential remedy pairs. We hypothesized that the most strongly associated pairs would be most clinically applicable. We validated each of the pairs against known medical practice and evaluated each for clinical feasibility. We then classified the pairs into categories that reflected the relationship with clinical practice, common practice, or potential harm, as follows: (1) substance is an FDA-approved or commonly used treatment for the symptom, (2) substance is not often used to treat the symptom but could be potentially useful given its pharmacological profile, (3) substance is a home or natural remedy for the symptom, (4) substance can cause the symptom, or (5) other or unclear (including cases where there was no clear connection between the potential remedy and the symptom; eg, gabapentin and fever).

Results

Entity Counts

Table 2 lists the top 20 most frequent substances and effects identified within the data set. As these are simply spans of text without semantic information, we can see duplicates (*fent* and *fentanyl*), slang terms (*dope* and *h*), and abbreviations (*rls*).

Table 2. Top 20 substances and effects extracted.

Entity	Count, n (%)[a]
Substances (N=2,823,606)	
dope	192,385 (6.81)
heroin	160,595 (5.69)
opiates	160,244 (5.68)
oxy	95,960 (3.40)
subs	87,006 (3.08)
opiate	80,067 (2.84)
methadone	76,477 (2.71)
kratom	68,541 (2.43)
fent	68,475 (2.43)
suboxone	65,011 (2.30)
h	62,492 (2.21)
sub	57,959 (2.05)
weed	51,937 (1.84)
morphine	49,114 (1.74)
benzos	37,208 (1.32)
tar	35,573 (1.26)
coke	35,355 (1.25)
xanax	34,765 (1.23)
fentanyl	30,361 (1.08)
codeine	29,050 (1.03)
Effects (N=479,289)	
pain	61,783 (12.89)
anxiety	30,995 (6.47)
depression	23,579 (4.92)
sleep	22,102 (4.61)
cravings	20,319 (4.24)
depressed	11,304 (2.36)
rls	8972 (1.87)
nausea	7618 (1.59)
craving	6315 (1.32)
insomnia	5645 (1.18)
mood	5600 (1.17)
puke	5551 (1.16)
seizures	5435 (1.13)
anxious	5343 (1.11)
tired	5159 (1.08)
sweating	5099 (1.06)
puking	5029 (1.05)
sickness	4931 (1.03)
sweat	4845 (1.01)
headache	4018 (0.84)

[a]The denominator for percentages is the total number of occurrences of all substances or effects, so percentage values do not sum to 100.

The model performs slightly better in predicting substances than effects (F_1 scores of 92.059 and 81.696). Recall (92.895 for substances and 83.768 for effects) is slightly higher than precision (91.237 for substances and 79.724 for effects). Both scores are reasonable given the model architecture used and the variability and quality of the input data in comparison with common benchmark NER tasks [43].

Deduplication and Expert Review

Both clustering and expert review greatly reduced the number of entities by removing misspellings and semantic duplicates.

Clustering reduced the number of unique substances by 98.07% (from 53,730 to 1037) and reduced the number of unique effects by 95.69% (from 13,790 to 594). Relative to the number of entities remaining after clustering, expert review reduced the number of unique substances by 55.8% (from 1037 to 458) and reduced the number of unique effects by 60.4% (from 594 to 235). Our final count of unique entities was 458 substances and 235 effects.

Effects were classified into 4 categories. Their frequencies are shown in Table 3.

Table 3. Frequency of effects by category (N=235).[a]

Effect	Values, n (%)
DSM-5[b] symptom of opioid withdrawal	17 (7.2)
Effect of opioid use	17 (7.2)
Not a DSM-5 symptom of opioid withdrawal or effect of opioid use	153 (65.1)
Other or unclear	48 (20.4)

[a]Percentages may not add up to 100 because of rounding.

[b]DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

Substances were classified into 71 pharmacological categories. The four most common were opioid (71/458, 15.5%); other (60/458, 13.1%); vitamin, supplement, or herb (42/458, 9.2%); and food or drink (29/458, 6.3%). Multimedia Appendix 2 includes frequencies for all 71 categories.

Network Analysis and Application

The Withdrawal Remedy Explorer application is publicly available [44]. It allows users to explore the associated substances or effects for a chosen entity. First, users can choose to view the substances or effects. Users then select an entity from a list of categorized substances or effects. This shows the ego network for the selected entity. The PPMI and edge count filters can then refine the network down to the strongest connections. For example, one could filter the network to the connections with the highest edge counts to view the most common substances associated with a symptom. Alternatively, filtering for the edges with the highest PPMI could uncover uncommon but noteworthy connections. Finally, connections with both high PPMI and high edge count are perhaps the most salient of all. Presenting the data as ego networks encourages users to identify substances or effects of interest before using the application. We considered this the most likely use case, as discussed further below. Figures 2 and 3 show example ego networks for a substance (acetaminophen) and effect (dehydration). Both ego networks are filtered for PPMI ≥1.5 and edge count ≥10.

Figure 2. Acetaminophen ego network filtered for positive pointwise mutual information ≥1.5 and edge count ≥10. DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; GERD: gastroesophageal reflux disease; RLS: restless leg syndrome.
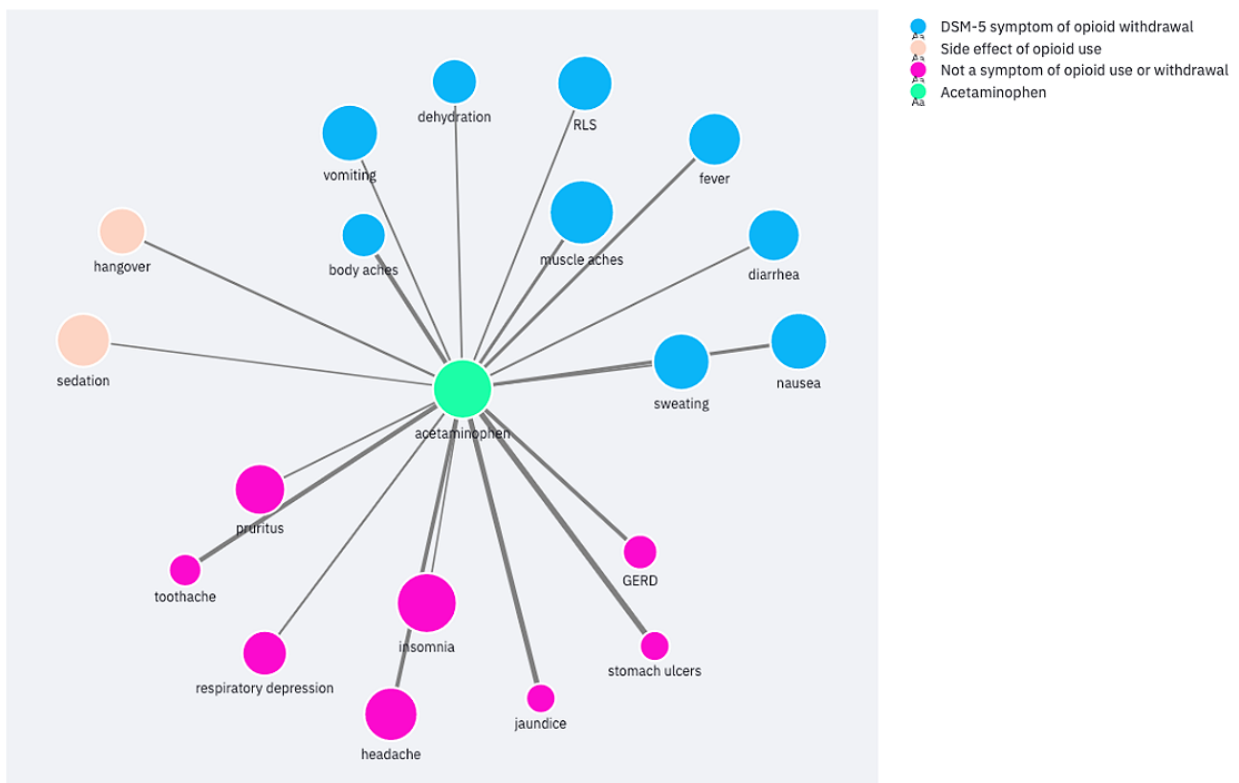
Figure 3. Dehydration ego network filtered for positive pointwise mutual information ≥1.5 and edge count ≥10. DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.
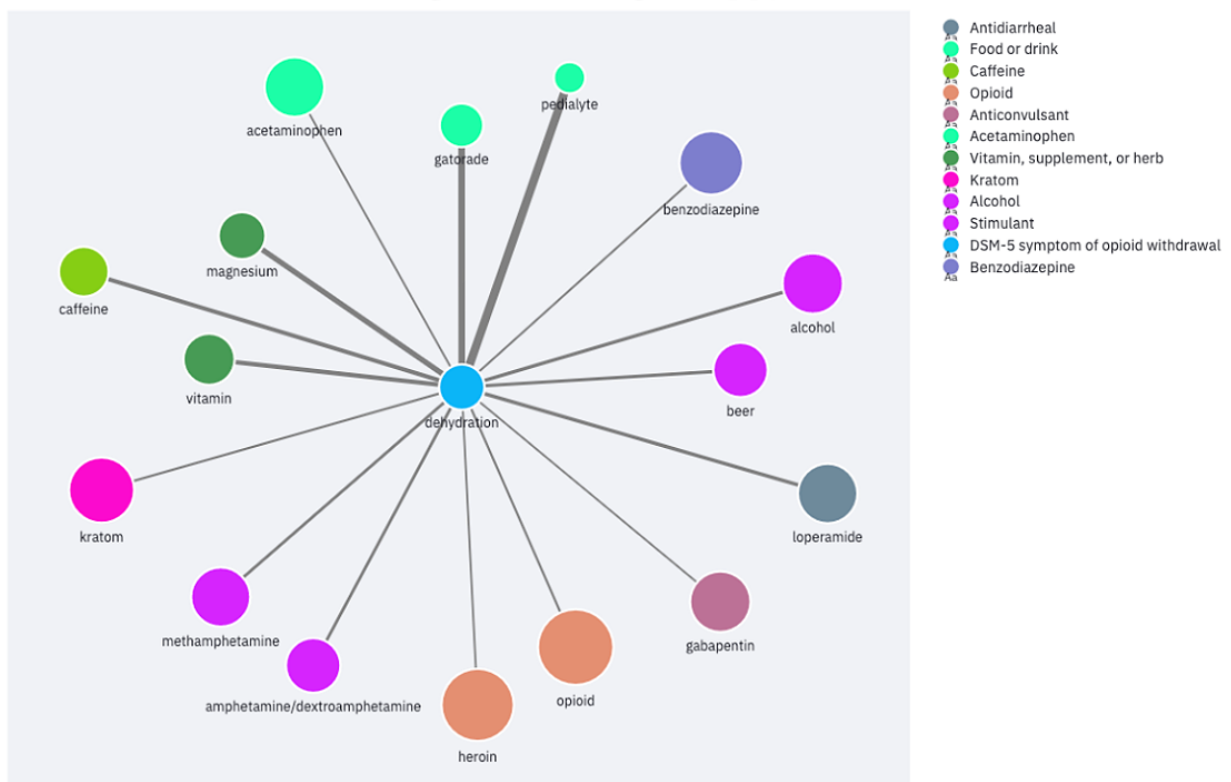
Validation Review of Pairs

In 64.6% (84/130) of the strongly associated symptom and potential remedy pairs, we considered the substance to be a potentially valid treatment for the symptom. This result provides evidence for the face validity of our methodology for extracting symptom–remedy pairs from Reddit. Furthermore, more strongly associated pairs were more likely to be potentially valid treatments. Of the 26 pairs in the top quintile by strength of association, 23 (88%) were potentially valid treatments for the symptom. Table 4 shows the category frequencies for the symptom and potential remedy pairs.

Table 4. Categorization of strongly associated symptom and potential remedy pairs (N=130).

Relationship	Substance-effect pairs, n (%)
Substance is an FDA[a]-approved or commonly used treatment for the symptom	54 (41.5)
Substance is not often used to treat the symptom but could be potentially useful given pharmacology	17 (13.1)
Substance is a home or natural remedy for the symptom	13 (10)
Substance can cause the symptom	7 (5.4)
Other or unclear (including pairs with no clear relationship)	39 (30)

[a]FDA: Food and Drug Administration.

Discussion

Principal Findings

In this study, we extracted information from Reddit forum posts into a database of opioid withdrawal symptoms and substances potentially used as remedies to alleviate them. Although the focus of the study was on withdrawal symptoms, we also extracted many other substances and effects associated with opioid use. We made a distinction between effects that are opioid withdrawal symptoms as defined in the DSM-5 and other effects. We validated the technical aspects of symptom and effect extraction by obtaining F_1 scores that were competitive with NER benchmarks.

We used PPMI to formalize the strength of substance-effect associations. We did not use any formal statistical tests or P values as the data themselves contain uncertainty and the population is not clearly defined. In fact, it is likely to change over time. However, the use of PPMI allowed us to compare the strength of associations with a *by chance* association and with each other.

We used PPMI to identify a list of 130 strongly associated symptom and potential remedy pairs in which the effect was a DSM-5 opioid withdrawal symptom and the substance was a potential remedy. We validated the associations between these symptoms and potential remedies based on expert review. We discovered that roughly two-thirds (84/130, 64.6%) of the potential remedies were common treatments, potentially useful treatments, or home or natural remedies. Therefore, we observed that, for DSM-5 withdrawal symptoms, the relationship between symptom and potential remedy validates well and that *common knowledge* is more prevalent than *urban myths*. We also developed a web application that allows researchers to explore these symptoms and potential remedies through visualization and to identify associations that may be strong enough to prompt further research.

Contributions

As demonstrated by the success of technical and expert review validations, our methodology has the potential to make significant contributions to ethnographic, clinical, and pharmacovigilance research. Specific areas include the following.

Understanding Potential Harm and Pharmacovigilance

Health care and public health stakeholders could benefit from knowing the types of self-medication and substitution practices that people engage in to help alleviate withdrawal symptoms, in part because those practices could lead to potential problems (eg, cardiovascular complications, medication contraindications, hospitalization, and even death). Analysis of social media discussions can rapidly inform prevention and harm-reduction activities related to new and potentially harmful beliefs and activities. This can enable stakeholders to monitor temporal trends and emerging fads. In fact, it was shown that increases in web-based posts about synthetic opioids preceded an increase in synthetic opioid death rates [45]. For example, in our study, we identified a strong association between antihistamines and restless leg syndrome (RLS). This could be a potential harm as the use of antihistamines as a remedy for RLS could in fact worsen RLS [46]. Our Withdrawal Remedy Explorer application allows one to screen for off-label use of many prescription drugs and thus assist in pharmacovigilance. Although such screening likely requires deeper follow-up with studying the actual posts, it provides a quick way to identify potential harms before they become more prevalent.

Identifying Home or Natural Remedies

Home or natural remedies have the potential to be inexpensive and effective measures against withdrawal symptoms. Such knowledge could help large numbers of struggling users, especially those who are not yet ready to receive treatment from a physician and who prefer the anonymity of 24/7 web-based communities to seek and share help [22]. For example, in our study, we identified a strong association between nausea and ginger, a common natural remedy [47]. Exploring associations between withdrawal symptoms and other herbs, vitamins, and supplements could help identify less well-known home or natural remedies.

Identifying Potentially Useful New Remedies

Successful off-label use of medications could provide leads to future clinical studies on withdrawal medications [48]. For example, in our study, we identified a strong association between gabapentin and body aches. Whether gabapentin is clinically effective for the body aches associated with opioid withdrawal is unknown; we were only able to identify 1 small study (N=32) investigating gabapentin as a treatment for opioid withdrawal [49]. Our results suggest that studies on gabapentin for opioid withdrawal may be fruitful. The identification of such remedies has not been the objective of this report, but we are partnering with clinicians and pharmacologists to identify such cases.

Understanding Patient Needs

Understanding patients' needs and issues that are of importance to them is critical for the development of better prevention and treatment programs. Web-based discussion forums and social media also provide mechanisms to inform and design better treatment and harm reduction programs. For example, future research could leverage the Withdrawal Remedy Explorer application to identify withdrawal symptoms, which, at least in these forums, are of the highest concern to people using opioids. In our study, 30% (39/130) of pairs fell into the *other or unclear* category. Although some of these pairs are likely the result of limitations to our methodology, this category provides opportunities for understanding users' beliefs and practices that go beyond *common knowledge*. Such information can lead to the discovery of new remedies or to the early identification of specific needs and potentially harmful practices.

Our approach can also be applicable to other substances of abuse, especially because currently there are no FDA-approved medications to treat cocaine, methamphetamine, and many other drug disorders. Uncovering and summarizing remedy practices for these disorders could provide at least temporary help in clinical treatment practice until treatment medications are developed and approved. Finally, by examining the symptom-related remedies, we acknowledge that some symptom–remedy pairs are quite common and are not specific to opioid withdrawal (eg, the use of melatonin for insomnia or acetaminophen for body aches). The detailed analysis of differences between remedies in opioid-specific discussions and in the general population is beyond the scope of this study.

Limitations and Future Work

Our study has several limitations. Foremost is the lack of causality in the associations between substances and effects. Without reading the posts, it is not possible to distinguish whether the effect was caused by the substance or whether the substance was used to alleviate the effect. Furthermore, we have not yet analyzed whether the remedies were helpful. An association could take many forms (eg, "gabapentin cured my body aches," "I tried gabapentin for body aches, but it didn't work," or "gabapentin gave me body aches"). However, the purpose of this study was to build and validate the foundations for such analyses. More detailed and in-depth analyses could be performed on any subset of the associations identified in the data. In future work, we plan to apply additional NLP methodologies and analysis to selected combinations of

substances and effects to identify whether the association was favorable.

We only focused on associations within the same sentence. This approach misses associations when the effect and substance are mentioned in different sentences (eg, "I get terrible body aches. Aspirin does nothing for me, but gabapentin helps"). In this study, we limited the approach to sentences for clarity and simplicity. With our success using single sentences, in future work we will focus on including associations spanning multiple sentences. This task is complicated by the necessity of defining the boundary of paragraphs and detangling multiple associations between multiple substances and effects across sentences. The ambiguity of free-flowing text in posts is a common challenge in the analysis of social media.

Since the time the data were collected and the models were developed, the field of NLP has progressed significantly because of the use of transformer models such as Bidirectional Encoder Representations from Transformers [50]. For general NER tasks, this has meant an improvement from an F_1 of 86 in 2017 to >90 in 2019. At the same time, large transformer models require significant specific computational resources [51] to train and deploy in comparison with more traditional methods and, simultaneously, focus in the field has been shifting from model-centric approaches (eg, hyperparameter tuning) to data-centric approaches (eg, higher-quality labeled data) as, in many scenarios, more benefit comes from data than architecture [52]. In summary, a limitation of our work is that we were unable to use the current state-of-the-art models, but future work will evaluate potential improvements from new model architectures and improved data quality.

In our expert review and validation processes, substances, effects, and substance-effect pairs were classified into categories according to the *face value* of the terms and without detailed assessment of the underlying post text. Different reviewers could have different opinions and interpretations, leading to variability in classifications. In future work, we plan to leverage the Withdrawal Remedy Explorer tool to seek comments and corrections from other experts.

As our study builds networks of extracted knowledge, in our future work we will consider developing a knowledge graph that would allow linking the extracted entities and their networks to external knowledge bases.

Despite these limitations, we identified and validated strong signals in the data. The discovery of valid withdrawal remedies encourages us to explore more nuanced aspects of the data. The limitations of this study point us to the next steps in our research.

Conclusions

In this study, we validated an approach to identify opioid withdrawal remedies from the web-based forum Reddit. We developed a pipeline to extract substances and effects from raw data, identified strong associations between withdrawal symptoms and potential remedies, and validated these associations. Our results demonstrate that social media and web-based forum discussions have the potential to help us understand how people treat withdrawal symptoms. This knowledge could help identify useful new treatments and

potential harms and public health concerns. We also developed the Withdrawal Remedy Explorer application to facilitate deeper analysis of these data and to seek input from other researchers, clinicians, and people with lived experience. Our approach could be generalized beyond Reddit and beyond the topic of opioid withdrawal. It could be applied to many other disease states where people self-manage their symptoms and discuss their experiences on the web.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed methods.
[DOCX File , 33 KB - formative_v6i3e33919_app1.docx]

Multimedia Appendix 2
Categorized substance frequencies.
[XLSX File (Microsoft Excel File), 12 KB - formative_v6i3e33919_app2.xlsx]

References

1. Kosten TR, George TP. The neurobiology of opioid dependence: implications for treatment. Sci Pract Perspect 2002;1(1):13-20 [FREE Full text] [doi: 10.1151/spp021113] [Medline: 18567959]
2. Blanco C, Volkow ND. Management of opioid use disorder in the USA: present status and future directions. Lancet 2019;393(10182):1760-1772. [doi: 10.1016/S0140-6736(18)33078-2] [Medline: 30878228]
3. Mannelli P, Peindl KS, Lee T, Bhatia KS, Wu LT. Buprenorphine-mediated transition from opioid agonist to antagonist treatment: state of the art and new perspectives. Curr Drug Abuse Rev 2012;5(1):52-63 [FREE Full text] [doi: 10.2174/1874473711205010052] [Medline: 22280332]
4. How do craving and withdrawal affect opioid relapse rates? Drug Rehab. URL: https://www.drugrehab.us/news/how-do-craving-and-withdrawal-affect-opioid-relapse-rates/ [accessed 2022-02-28]
5. Unnithan S, Gossop M, Strang J. Factors associated with relapse among opiate addicts in an out-patient detoxification programme. Br J Psychiatry 1992;161:654-657. [doi: 10.1192/bjp.161.5.654] [Medline: 1422615]
6. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5. 5th edition. Washington, D.C: American Psychiatric Association; 2013:547-548.
7. Kosten TR, Baxter LE. Review article: effective management of opioid withdrawal symptoms: a gateway to opioid dependence treatment. Am J Addict 2019;28(2):55-62 [FREE Full text] [doi: 10.1111/ajad.12862] [Medline: 30701615]
8. Volkow ND, Collins FS. The role of science in addressing the opioid crisis. N Engl J Med 2017;377(4):391-394. [doi: 10.1056/NEJMsr1706626] [Medline: 28564549]
9. Doughty B, Morgenson D, Brooks T. Lofexidine: a newly FDA-approved, nonopioid treatment for opioid withdrawal. Ann Pharmacother 2019;53(7):746-753. [doi: 10.1177/1060028019828954] [Medline: 30724094]
10. FDA news release - FDA approves the first non-opioid treatment for management of opioid withdrawal symptoms in adults. U.S. Food and Drug Administration. 2018. URL: https://www.fda.gov/news-events/press-announcements/fda-approves-first-non-opioid-treatment-management-opioid-withdrawal-symptoms-adults [accessed 2022-02-28]
11. Kuszmaul AK, Palmer EC, Frederick EK. Lofexidine versus clonidine for mitigation of opioid withdrawal symptoms: a systematic review. J Am Pharm Assoc (2003) 2020;60(1):145-152. [doi: 10.1016/j.japh.2019.10.004] [Medline: 31791720]
12. Generali JA, Cada DJ. Baclofen: opiate withdrawal. Hosp Pharm 2008;43(9):716-717 [FREE Full text] [doi: 10.1310/hpj4309-716]
13. Peckham AM, Evoy KE, Ochs L, Covvey JR. Gabapentin for off-label use: evidence-based or cause for concern? Subst Abuse 2018;12:1178221818801311 [FREE Full text] [doi: 10.1177/1178221818801311] [Medline: 30262984]
14. Gowing L, Farrell M, Ali R, White JM. Alpha2-adrenergic agonists for the management of opioid withdrawal. Cochrane Database Syst Rev 2016;2016(5):CD002024 [FREE Full text] [doi: 10.1002/14651858.CD002024.pub5] [Medline: 27140827]
15. Srivastava AB, Mariani JJ, Levin FR. New directions in the treatment of opioid withdrawal. Lancet 2020;395(10241):1938-1948 [FREE Full text] [doi: 10.1016/S0140-6736(20)30852-7] [Medline: 32563380]

16.   Opiates. Reddit. URL: https://www.reddit.com/r/opiates/ [accessed 2022-02-28]
17.   Alfonso III F. How a Reddit forum has become a lifeline to opioid addicts in the US. The Guardian. 2017. URL: https:/
        /www.theguardian.com/society/2017/jul/19/opioid-addiction-reddit-fentanyl-appalachia [accessed 2022-02-28]
18.   You are not alone in this fight. Reddit. URL: https://www.reddit.com/r/OpiatesRecovery/ [accessed 2022-02-28]
19.   Chancellor S, Nitzburg G, Hu A, Zampieri F, De Choudhury M. Discovering alternative treatments for opioid use recovery
        using social media. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented
        at: CHI '19; May 4-9, 2019; Glasgow, UK p. 1-15. [doi: 10.1145/3290605.3300354]
20.   Daniulaityte R, Falck R, Carlson RG. Sources of pharmaceutical opioids for non-medical use among young adults. J
        Psychoactive Drugs 2014;46(3):198-207 [FREE Full text] [doi: 10.1080/02791072.2014.916833] [Medline: 25052878]
21.   Mitchell JT, Sweitzer MM, Tunno AM, Kollins SH, McClernon FJ. "I Use Weed for My ADHD": a qualitative analysis
        of online forum discussions on cannabis use and ADHD. PLoS One 2016;11(5):e0156614 [FREE Full text] [doi:
        10.1371/journal.pone.0156614] [Medline: 27227537]
22.   D'Agostino AR, Optican AR, Sowles SJ, Krauss MJ, Escobar Lee K, Cavazos-Rehg PA. Social networking online to recover
        from opioid use disorder: a study of community interactions. Drug Alcohol Depend 2017;181:5-10 [FREE Full text] [doi:
        10.1016/j.drugalcdep.2017.09.010] [Medline: 29024875]
23.   Paul MJ, Chisolm MS, Johnson MW, Vandrey RG, Dredze M. Assessing the validity of online drug forums as a source for
        estimating demographic and temporal trends in drug use. J Addict Med 2016;10(5):324-330. [doi:
        10.1097/ADM.0000000000000238] [Medline: 27466069]
24.   Bowen DA, O'Donnell J, Sumner SA. Increases in online posts about synthetic opioids preceding increases in synthetic
        opioid death rates: a retrospective observational study. J Gen Intern Med 2019;34(12):2702-2704 [FREE Full text] [doi:
        10.1007/s11606-019-05255-5] [Medline: 31468343]
25.   Elbattah M, Arnaud É, Gignon M, Dequen G. The role of text analytics in healthcare: a review of recent developments and
        applications. In: Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies.
        2021 Presented at: BIOSTEC '21; February 11-13, 2021; Online p. 825-832. [doi: 10.5220/0010414508250832]
26.   Baumgartner J, Zannettou S, Keegan B, Squire M, Blackburn J. The pushshift Reddit dataset. In: Proceedings of the 14th
        International AAAI Conference on Web and Social Media. 2020 Presented at: ICWSM '08; March 30-April 2, 2008; Seattle,
        WA p. 830-839 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/7347
27.   Baumgartner P. RTIInternational/PushshiftRedditDistiller: initial release. Meta. 2020. URL: https://www.meta.org/software/
        rtiinternationalpushshiftredditdistiller-initial/software-165049 [accessed 2022-02-28]
28.   Mikolov T, Corrado GS, Checn K, Dean J. Efficient estimation of word representations in vector space. In: Proceedings
        of the International Conference on Learning Representations. 2013 Presented at: ICLR '13; April 30-May 3, 2013; Vancouver,
        Canada.
29.   Rehurek R, Sokja P. Software framework for topic modelling with large corpora. In: Proceedings of Workshop New
        Challenges for NLP Frameworks. 2010 Presented at: LREC '20; May 22, 2010; Valetta, Malta p. 46-50.
30.   Neumann M, King D, Beltagy I, Ammar W. Scispacy: fast and robust models for biomedical natural language processing.
        In: Proceedings of the 18th BioNLP Workshop and Shared Task. 2019 Presented at: BioNLP@ACL '19; August 1, 2019;
        Florence, Italy p. 319-327. [doi: 10.18653/v1/w19-5034]
31.   Prodigy. URL: https://prodi.gy/ [accessed 2022-02-28]
32.   Campello RJ, Moulavi D, Sander J. Density-based clustering based on hierarchical density estimates. In: Pacific-Asia
        Conference on Knowledge Discovery and Data Mining. 2013 Presented at: PAKKD '13; April, 14-17, 2013; Gold Coast,
        Australia p. 160-172 URL: https://doi.org/10.1007/978-3-642-37456-2_14 [doi: 10.1007/978-3-642-37456-2_14]
33.   McInnes L, Healy J, Astels S. hdbscan: hierarchical density based clustering. J Open Source Soft 2017;2(11):205 [FREE
        Full text] [doi: 10.21105/joss.00205]
34.   The fastest way to build and share data apps. Streamlit. URL: https://streamlit.io/ [accessed 2022-02-28]
35.   Collaborative data science. Plotly Technologies. 2015. URL: https://plot.ly [accessed 2022-02-28]
36.   Withdrawal remedy explorer. GitHub. URL: https://github.com/RTIInternational/withdrawal-remedy-explorer [accessed
        2022-02-28]
37.   Montani I, Honnibal M, Van Landeghem S, Peters H, Boyd A, Samsonov M, Roman, Explosion Bot, GregDuppin,
        jeannefukumaru, et al. zenodo. 2020. URL: https://zenodo.org/record/3897194 [accessed 2022-02-28]
38.   Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks
        from overfitting. J Mach Learn Res 2014;15(56):1929-1958 [FREE Full text]
39.   Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching word vectors with subword information. Trans Assoc Comput
        Linguist 2017;5:135-146. [doi: 10.1162/tacl_a_00051]
40.   Balsamo D, Bajardi P, Salomone A, Schifanella R. Patterns of routes of administration and drug tampering for nonmedical
        opioid consumption: data mining and content analysis of Reddit discussions. J Med Internet Res 2021;23(1):e21212 [FREE
        Full text] [doi: 10.2196/21212] [Medline: 33393910]
41.   Aji S, Kaimal R. Document summarization using positive pointwise mutual information. Int J Comput Sci Inf Technol
        2012;4(2):47-55 [FREE Full text] [doi: 10.5121/ijcsit.2012.4204]

42. Everett M, Borgatti SP. Ego network betweenness. Soc Netw 2005;27(1):31-38 [FREE Full text] [doi: 10.1016/j.socnet.2004.11.007]

43. Named entity recognition on Ontonotes v5 (English). Papers with Code. URL: https://paperswithcode.com/sota/named-entity-recognition-ner-on-ontonotes-v5 [accessed 2022-02-28]

44. Withdrawal remedy explorer. RTI International. 2021. URL: https://share.streamlit.io/rtiinternational/withdrawal-remedy-explorer [accessed 2022-02-28]

45. Eban K. OxyContin: purdue pharma's painful medicine. Fortune. 2011. URL: http://fortune.com/2011/11/09/oxycontin-purdue-pharmas-painful-medicine/ [accessed 2022-02-28]

46. Bliwise DL, Zhang RH, Kutner NG. Medications associated with restless legs syndrome: a case-control study in the US Renal Data System (USRDS). Sleep Med 2014;15(10):1241-1245 [FREE Full text] [doi: 10.1016/j.sleep.2014.05.011] [Medline: 25156752]

47. Ernst E, Pittler MH. Efficacy of ginger for nausea and vomiting: a systematic review of randomized clinical trials. Br J Anaesth 2000;84(3):367-371 [FREE Full text] [doi: 10.1093/oxfordjournals.bja.a013442] [Medline: 10793599]

48. Krupitsky EM, Ilyuk RD, Mikhailov AD, Kazankov KA, Rybakova KB, Skurat EP, et al. A randomized single blind study of the efficacy of pregabalin in the treatment of opioid withdrawal syndrome. Zh Nevrol Psikhiatr Im S S Korsakova 2016;116(7):29-36. [doi: 10.17116/jnevro20161167129-36] [Medline: 27500873]

49. Medical University of South Carolina. Discontinuation from chronic opioid therapy for pain using a buprenorphine taper. ClinicalTrials. 2016. URL: https://doi.org/https://clinicaltrials.gov/ct2/show/NCT02737826 [accessed 2022-02-28]

50. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2018 Presented at: NAACL '18; June 1-6, 2018; New Orleans, LA URL: http://arxiv.org/abs/1810.04805

51. Strubell E, Ganesh A, McCallum A. Energy and policy considerations for modern deep learning research. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2019 Presented at: ICAI '19; February 2-9, 2021; Onine URL: http://arxiv.org/abs/1906.02243 [doi: 10.18653/v1/p19-1355]

52. Kaplan J, McCandish S, Henighan T, Brown TB, Chess B, Child R, et al. Scaling laws for neural language models. arXiv. Preprint posted online on January 23, 2020. [FREE Full text]

Abbreviations

DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
FDA: Food and Drug Administration
NER: named entity recognition
NLP: natural language processing
PPMI: positive pointwise mutual information
RLS: restless leg syndrome

Original Paper

eHealth Literacy in a Sample of South Asian Adults in Edmonton, Alberta, Canada: Subanalysis of a 2014 Community-Based Survey

Mark J Makowsky[1], BSP, PharmD; Shahnaz Davachi[2], RD, PhD; Charlotte A Jones[3], MD, PhD

[1]Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton, AB, Canada

[2]Primary Health Care, Alberta Health Services, Calgary, AB, Canada

[3]Faculty of Medicine, Southern Medical Program, University of British Columbia Okanagan Campus, Kelowna, BC, Canada

Corresponding Author:
Mark J Makowsky, BSP, PharmD
Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta
2-142E Katz Group - Rexall Centre for Pharmacy & Health Research
11361 87 Avenue
Edmonton, AB, T6G 2E1
Canada
Phone: 1 7804921735
Fax: 1 7804921217
Email: makowsky@ualberta.ca

Abstract

Background: Digital health interventions are efficient and flexible methods for enhancing the prevention and management of cardiovascular disease and type 2 diabetes. However, little is known about the characteristics associated with eHealth literacy in the Canadian South Asian population.

Objective: The aim of this study is to describe perceived eHealth literacy and explore the extent to which it is associated with sociodemographic, health status, and technology use variables in a subset of South Asian Canadians.

Methods: We analyzed data from the e-Patient Project survey, a mixed-mode cross-sectional survey that occurred in 2014. The eHealth Literacy Scale (eHEALS) was used to measure eHealth literacy in a convenience sample of 511 English- or Punjabi-speaking South Asian adults recruited from a community pharmacy, a family physician office, and community events in Edmonton, Alberta. Multivariable quantile regression was used to explore variables associated with eHealth literacy.

Results: The analysis was restricted to 301 internet users (mean age 39.9, SD 14.8 years; 166/301, 55.1% female) who provided responses to all 8 eHEALS questions and complete demographic information. The mean overall eHEALS score was 29.3 (SD 6.8) out of 40, and 71.4% (215/301) agreed to at least 5 out of the 8 eHEALS items. The eHEALS item with the lowest level of agreement was "I can tell high-quality health resources from low-quality health resources on the internet" (182/301, 60.5%). Although there were statistically significant differences in eHEALS scores according to age, educational achievement, language preference, and the presence of chronic medical conditions, multivariable regression analysis indicated that language preference was the only variable independently associated with eHealth literacy (coefficient –6.0, 95% CI –9.61 to –2.39).

Conclusions: In our sample of South Asian Canadian internet users, preference for written health information in languages other than English was associated with lower eHealth literacy. Opportunities exist to improve eHealth literacy using culturally and linguistically tailored interventions.

KEYWORDS

XSL•FO
RenderX

Introduction

Background

Digitally engaged patients are those who have taken up new digital media technologies in their own medical care and preventative health efforts [1]. These eHealth technologies typically support self-care and include internet-based resources where lay people can seek information about health, web-based communities supported by social networking sites, mobile health (mHealth)—the practice of medicine and public health with support from mobile devices, telemedicine and remote patient monitoring where patients communicate with health care providers via technology instead of face-to-face communication, and others [1-3]. Emerging evidence suggests that eHealth-and mHealth-based interventions can improve the prevention and management of chronic health conditions [4,5]. For example, systematic reviews suggest the potential for the improvement of cardiovascular lifestyle-related risk factors [6] including weight loss [7-9], physical activity [10,11], and management of diabetes [12-14].

To optimally engage with and have equitable access to a digital health care environment, health care consumers must have an understanding of their condition as well as the skills to effectively use electronic resources [15-17]. This is referred to as eHealth literacy, which is defined as the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to address or solve a health problem [18]. There are six main domains of eHealth literacy: health, computers, media, science, and information literacy, as well as traditional literacy and numeracy [18]. eHealth literacy has taken on greater importance as the COVID-19 pandemic has shifted primary care from office visits to telephone or video care in the Canadian health care system [19] and resulted in a broader *infodemic* of medical misinformation [20,21]. Assessing eHealth literacy is important because it has been associated with positive outcomes from internet searches in health knowledge and information gathering, self-management of health needs and health behaviors, and interactions with physicians [22-25]. eHealth literacy is modifiable, and several studies have shown that it can be improved in older individuals and those with chronic health conditions [15,26,27].

South Asian Canadians are a large visible minority group [28] who face a high prevalence of cardiovascular disease and diabetes, both of which are conditions for which self-management are required [29-31]. Expanded access to and use of culturally appropriate eHealth strategies may assist South Asian Canadians in addressing documented gaps in risk factors and diabetes control [30] by addressing documented barriers [32,33] such as language, sociocultural factors, misconceptions around diet and physical activity, lack of access to culturally tailored diet counseling, and compliance with pharmacotherapy.

Whether South Asian Canadians have adequate eHealth literacy to effectively use eHealth to improve their health is uncertain as are the demographic and technology use determinants of eHealth literacy in this community. Although eHealth strategies may support improving both the prevention and management

of cardiovascular disease and type 2 diabetes, a systematic review of eHealth literacy suggests that underserved populations may have lower levels of competence in the 6 core domains of eHealth literacy, decreased access to health infrastructure, and technology [34]. For example, research conducted by Statistics Canada suggests that immigrants to Canada originating from non–English- and non–French-speaking countries score below the national average in health literacy. On the basis of 2003 survey data, fewer immigrants (24%) than nonimmigrants (44%) had requisite levels of health literacy [35]. Research on Danish immigrants has suggested that descendants and immigrant women have lower levels of eHealth literacy and health literacy than women of Danish origin [36]. In contrast, data on immigrants to Israel suggest that language barriers because of immigration, negatively impact health literacy but have no impact on eHealth literacy [37]. Research on primary digital divides in internet access by race or ethnicity has been conflicting, with some suggesting no difference in digital use divides [38], whereas others suggest that they do exist [39]. Demographic and socioeconomic variables associated with eHealth literacy in the literature are inconsistent and are suggested to be population-dependent [17].

Objectives

Owing to a lack of information on digital device use and eHealth literacy in South Asian Canadians, we conducted the South Asian e-Patient Project, a survey conducted in a convenience sample of 831 South Asian Canadians in Edmonton, Alberta in 2014 [40]. We previously reported that engaging members of this community via eHealth interventions is feasible. However, we found evidence of digital divides in the use of the internet, digital devices, and apps for health purposes by language preference, education, age, gender, confidence in filling out medical forms, and the number of years lived in Canada. Further description and examination of the characteristics associated with eHealth literacy in members of the Canadian South Asian Community is important because this information could be used to identify levels of readiness for eHealth, to inform and justify the development of tailored solutions to overcome identified gaps in the required skills to effectively apply web-based health information and other mHealth interventions and to inform and assist health care providers to optimally engage individuals in remote internet-based care and with existing web-based health information resources. Therefore, the objectives of this study are to (1) describe levels of eHealth literacy among a subset of South Asian adult internet users who were invited to complete the eHealth Literacy Scale (eHEALS) as part of the community-based e-Patient Project survey and (2) explore sociodemographic, health status, and technology use variables and their association with eHealth literacy among survey respondents.

Methods

e-Patient Project Study Design

The design as well as the results on the prevalence and variables associated with internet and digital device use for health purposes from the e-Patient Project survey have been published previously [40]. Briefly, the survey was an anonymous,

mixed-mode survey conducted with a convenience sample of English- or Punjabi-speaking South Asian adults between May 18 and August 31, 2014. We used a community-based approach and worked in partnership with 13 faith-based, cultural, community, and health care organizations in Edmonton, Alberta. The survey was designed to evaluate levels of engagement with the internet, digital device ownership, use of health and fitness apps, health information–seeking practices, preferences for delivery modalities for future eHealth interventions, and eHealth literacy. (Multimedia Appendix 1) The survey was primarily a computer-assisted in-person interview conducted at faith-based gathering places, health care settings, community centers, and events using the Qualtrics (Qualtrics Corporation) web-based survey platform. Most questions were adopted from existing survey instruments, including the Pew Research Center's Internet & American Life Project 2012 Health survey [41-43] and the 2012 Statistics Canada Canadian Internet Use survey [44]. The survey was pilot-tested with 19 individuals from the target communities.

eHealth Literacy Assessment—Inclusion Criteria, Recruitment, and Data Collection

A subset of the survey participants was invited to complete the eHealth literacy assessment. Owing to the time required to complete the survey, this subset included all individuals recruited in person at the participating community pharmacy, family physician practice, a large community event, and 1 community center, as well as those who completed the survey via the web. Individuals recruited at faith-based gatherings were excluded from the eHealth literacy assessment.

Potential respondents were notified about the study via personal invitations by pharmacy or clinic staff, announcements at community events, and posters. We recruited consecutive attendees from the community pharmacy and family physician office sites. Individuals were then approached by trained, bilingual English- and Punjabi-speaking research assistants or community volunteers. They were presented with the e-Patient Project survey information letter and were asked if they would like to participate. The agreement and completion of the survey implied participants' consent.

Research staff administered the survey, including the eHealth literacy assessment, via computer-assisted personal interviews in English or Punjabi, according to respondent preferences. One-on-one interviews using paper surveys were conducted at the community event before the tablet computers became available. A web-based version was also offered to those who preferred to complete the survey on their own time. Respondents who completed the survey were offered a reusable shopping bag and the opportunity to enter a draw for a tablet computer or various gift cards as incentives.

Measurement of eHealth Literacy

eHealth literacy was measured using the 8-item eHEALS scale, which is the most commonly used validated measure of eHealth literacy [45,46]. The eHEALS measures the concept of eHealth literacy, defined as a set of skills required to effectively engage information technology for health and has shown high levels of internal consistency and test-retest reliability [46]. Each of

the 8 items is rated on a 5-point Likert scale (1=strongly disagree to 5=strongly agree). The overall score ranges from 8 to 40, with higher scores suggesting higher eHealth literacy. In addition, 2 supplemental eHEALS items, which do not contribute to the overall score, were also included before the 8 items to measure the perceived usefulness of the internet to help make health decisions and the perceived importance of being able to access health resources on the internet.

We created a Punjabi version of the survey and the eHEALS instrument according to the World Health Organization guidelines for translation and adaptation of instruments [47] (Multimedia Appendix 2). One Punjabi-speaking translator with a medical background who was also fluent in English conducted a forward translation from English to Punjabi. Emphasis was placed on conceptual rather than literal translation. A panel of 2 bilingual community member reviewers further reviewed and identified inadequate expressions and concepts in the translated versions. The back translation was conducted by a separate translator who was fluent in both English and Punjabi. The translated eHEALS was included in the pilot test, and translation discrepancies were discussed and addressed by the project team. The internal consistency for the 8 item eHEALS for the 301 respondents was Cronbach α=.950, and principal components analysis produced a single factor solution with factor loadings from 0.68 to 0.80 among the 8 items, with eigenvalue=5.95 and 74.4% of the variance explained (Multimedia Appendix 3).

Measurement of Sociodemographic and Internet Use Variables

Demographic factors included age, sex, education, marital status, duration of time lived in Canada, and South Asian community affiliation. Individuals who answered affirmatively to either *Do you go online at least occasionally* or *Do you send or receive email at least occasionally*? were characterized as internet users. Language preference was determined by asking *In what language would you prefer to receive written health information* and the categories were collapsed into includes English or does not include English. We estimated health literacy using the question *How confident are you filling out medical forms by yourself?* that was effective in detecting inadequate health literacy as measured using the Short Test of Functional Health Literacy in Adults [48]. Individuals who indicated being likely or very likely to use at least one of six different modes of eHealth interventions in the next 12 months were deemed to be interested in eHealth interventions.

Statistical Analysis

We limited the analysis to internet user respondents who provided complete responses to all 8 eHEALS questions and complete demographic information for the planned multivariate analysis. Surveys with other missing data were included in the analysis, and descriptive statistics for categorical variables were depicted as proportions of total cases, including those with missing data. SPSS (version 28; IBM Corp) was used to compute descriptive statistics for the demographic characteristics, technology use variables, and eHEALS scores. Summary eHEALS scores were calculated by summing the responses to 8 eHEALS questions. We also explored the proportion of individuals who agreed to at least five out of the eight eHEALS

items, an indicator of adequate eHealth literacy [49], and the number of participants scoring 26 or more, which has been considered an indicator of high eHealth literacy [17,50]. We explored bivariate differences in eHEALS scores across demographic, health status, and technology use outcome variables using the 2-tailed Welch t test or Welch F test as appropriate [51-57]. Statistical significance was established at $P<.05$ and Bonferroni correction was used for post hoc tests.

Multivariable quantile regression was performed using R (version 3.1.3; The R Foundation) to assess the effect of each demographic variable on the eHEALS score while controlling for the effect of others. This analysis was restricted to internet users and explored both demographic and internet use predictors. Quantile regression through the median was performed instead of mean regression [58] because regular regression residuals differed substantially from normality and violated the constant error variance assumption as well. Log and square root transformations also did not work. Quantile regression models conditional quantiles instead of conditional means. Here, we model the conditional median (the 0.5th quantile). It does not assume normality and constant error variance and is robust against outliers. The CIs and P values reported were based on bootstrap SEs with 1000 repetitions.

Ethics Approval

The Health Research Ethics Board at the University of Alberta (Pro00038210) approved this study.

Results

Participant Flow

A total of 511 individuals were invited to complete the eHealth Literacy Assessment. Although we originally invited both internet users and noninternet users to complete the eHEALS assessment, given that it is more meaningful to investigate eHealth literacy in internet users, in this paper, we focus on reporting the results in the n=301 internet users who provided data for all 8 HEALS items and demographic characteristics. (Figure 1). Findings from the 83 noninternet user eHEALS responders are shown in Multimedia Appendix 4.

Figure 1. Flow diagram of the eHealth Literacy Scale assessment in the e-Patient Project Survey, Edmonton, Alberta, 2014.
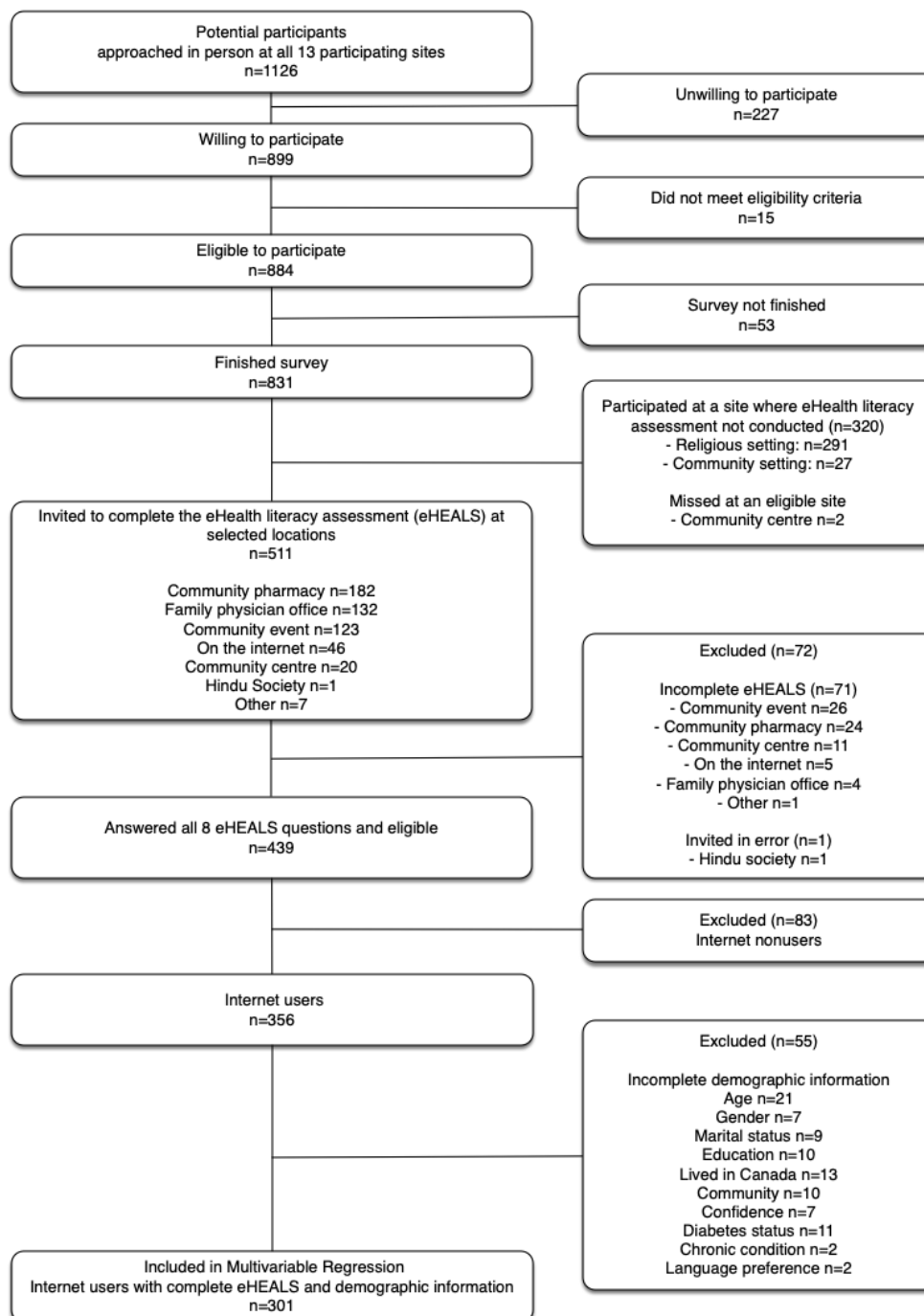
Participant Characteristics

Table 1 shows the demographic characteristics and the mean eHEALS scores of the study sample. The mean age was 39.9 (SD 14.8) years and just over half of the respondents were female (166/301, 55.1%). Most respondents were married (235/301, 78.1%); had college, university, or higher education (238/301, 79.1%); and self-identified as Sikh (212/301, 70.4%). A total of 29.2% (88/301) lived in Canada for <5 years. Most

respondents (260/301, 86.4%) indicated that English was their preferred language for written health information. Overall, 81.8% (246/301) indicated that their health was at least good, 53.2% (160/301) had at least 1 of the 8 chronic health conditions we inquired about, and only 1.7% (5/301) reported not being confident at all in filling out medical forms by themselves. Most respondents (194/301, 64.5%) were recruited from the participating community pharmacy and family physician office.

Table 1. Demographic characteristics of respondents and mean overall eHEALS[a] scores among South Asian study participants by demographic characteristics.

Demographic	Values (n=301), n (%)	eHEALS score, mean (SD)	Difference in eHEALS scores (95% CI)	P value[b]
Age (years)				.009[c]
<65	275 (91.4)	29.68 (6.55)	Referent	
≥65	26 (8.6)	24.92 (8.40)	–4.76 (–8.23 to –1.29)	
Sex				.14[c]
Male	135 (44.9)	28.61 (7.79)	Referent	
Female	166 (55.1)	29.81 (5.92)	1.21 (–2.81 to 0.40)	
Marital status				.12[c]
Not married	66 (21.9)	30.55 (7.65)	Referent	
Married	235 (78.1)	28.91 (6.57)	–1.63 (–3.69 to 0.42)	
Education				.01[d]
<High school	3 (1)	15.67 (12.42)	Referent	N/A[e]
High School	60 (19.9)	25.27 (7.91)	9.60 (–31.32 to 50.52)	.51[f]
≥College	238 (79.1)	30.45 (5.88)	14.79 (–27.27 to 56.84)	.30[f]
Lived in Canada (years)				.57[e]
>5	213 (70.8)	29.41 (7.07)	Referent	
0-5	88 (29.2)	28.94 (6.29)	–0.47 (–2.10 to 1.66)	
South Asian community				.16[d]
Sikh	212 (70.4)	28.83 (7.28)	Referent	N/A
Hindu	42 (14)	30.40 (5.54)	1.57 (–0.79 to 3.94)	.26[f]
Other	47 (15.6)	30.26 (5.64)	1.43 (–0.87 to 3.72)	.31[f]
Confidence in filling out medical forms				.35[g]
>Not at all	296 (98.3)	29.36 (6.68)	Referent	
Not at all	5 (1.7)	23.80 (13.55)	–5.57 (–19.06 to 7.83)	
Health status				.83[h]
Excellent	43 (14.3)	28.23 (9.18)	Referent	
Very good	70 (23.3)	29.81 (6.57)	1.58 (–2.92 to 6.08)	
Good	133 (44.2)	29.50 (6.21)	1.27 (–2.96 to 5.50)	
Fair	41 (13.6)	28.61 (6.95)	0.38 (–4.57 to 5.32)	
Poor	14 (4.7)	29.50 (5.63)	1.27 (–4.62 to 7.16)	
Diabetes				.58[c]
No	259 (86)	29.36 (6.94)	Referent	
Yes	42 (14)	28.76 (6.25)	–0.59 (–2.71 to 1.52)	
High blood pressure				.02[c]
No	241 (80.1)	29.9 (6.30)	Referent	
Yes	55 (18.3)	26.95 (8.35)	–2.95 (–5.34 to –0.56)	
Heart disease				.51[c]
No	279 (92.7)	29.5 (6.69)	Referent	
Yes	12 (4)	27.5 (10.05)	–2.0 (–8.42 to 4.42)	

XSL•FO

RenderX

Demographic	Values (n=301), n (%)	eHEALS score, mean (SD)	Difference in eHEALS scores (95% CI)	P value[b]
Lung conditions				.68[c]
No	274 (91)	29.29 (6.87)	Referent	
Yes	19 (6.3)	29.95 (6.50)	0.66 (–2.57 to 3.88)	
Arthritis				.14[c]
No	261 (86.7)	29.49 (6.76)	Referent	
Yes	31 (10.3)	27.29 (7.70)	–2.20 (–5.13 to 0.74)	
Cancer				.05[c]
No	283 (94)	29.53 (6.73)	Referent	
Yes	11 (3.7)	23.82 (8.54)	–5.71 (–11.48 to 0.06)	
Other chronic condition				.046[c]
No	244 (81.1)	29.72 (6.83)	Referent	
Yes	45 (15)	27.49 (6.75)	–2.23 (–4.42 to –0.04)	
High cholesterol				.16[c]
No	193 (64.1)	30.15 (5.67)	Referent	
Yes	41 (13.6)	28.24 (8.16)	–1.90 (–4.59 to 0.79)	
≥1 Chronic condition[i]				.08[c]
No	160 (53.2)	29.93 (6.42)	Referent	
Yes	141 (46.8)	28.52 (7.24)	–1.41 (–2.97 to 0.16)	
Language preference				<.001[f]
English	260 (86.4)	30.38 (5.86)	Referent	
Not English	41 (13.6)	22.22 (8.32)	–8.17 (–10.80 to –5.45)	
Location of recruitment				.03[h]
Community setting	70 (23.3)	27.21 (8.21)	Referent	N/A
Health setting	194 (64.5)	30.07 (6.31)	2.86 (0.29 to 5.43)	.03[f]
Web-based	37 (12.3)	28.97 (5.93)	1.76 (–1.53 to 5.05)	.42[f]

[a]eHEALS: eHealth Literacy Scale.

[b]*P* value is for the comparison of eHEALS scores.

[c]Welch *t* test.

[d]Welch *F* test on trimmed mean and Winsorized variance as data violated the constant variance assumption and were far from normal.

[e]N/A: not applicable.

[f]Pairwise comparison using the Games-Howell post hoc method; a *P* value of 0.5/number of comparisons was considered statistically significant.

[g]The bootstrap method with 1000 repetitions was used, as the data differed substantially from the normal distribution. For the bootstrap method, bias-corrected and accelerated CIs were reported.

[h]The Welch *F* test was used as the data violated the equal variance assumption of the analysis of variance.

[i]Chronic conditions included diabetes or sugar disease, high blood pressure, heart disease (eg, angina, heart attack, or stroke), lung conditions (eg, asthma or bronchitis), arthritis, cancer, high cholesterol, or other chronic conditions treated with daily medication.

Overall eHEALS Scores and Patterns of Item Responses

Total scores on the eHEALS ranged from 8 to 40 and were negatively skewed with a mean of 29.27 (SD 6.84), median of 31, and IQR of 27 to 32. A total of 2.3% (7/301) participants had the worst possible score, whereas 4.7% (14/301) of participants had the best possible score. Over three quarters of respondents (253/299, 84.6%) felt it is important to be able to access health resources on the internet and that the internet is useful in helping make decisions about their health (234/298, 78.5%). The proportion of respondents who agreed with each eHEALS item is shown in Figure 2. Almost three quarters of respondents (215/301, 71.4%) had adequate health literacy (ie, agreed to at least 5 out of the 8 eHEALS items) and 78.1% (235/301) had eHEALS scores ≥26. The 2 items with the lowest levels of agreement were for *I can tell high-quality health resources from low-quality health resources on the internet*

(182/301, 60.5%) and *I feel confident in using information from*   *the internet to make health decisions* (191/301, 63.5%).

Figure 2. Frequency of responses to 8-item eHealth Literacy Scale among South Asian study participants (n=301 internet users). Data for strongly disagree and disagree were collapsed as was agree and strongly agree.
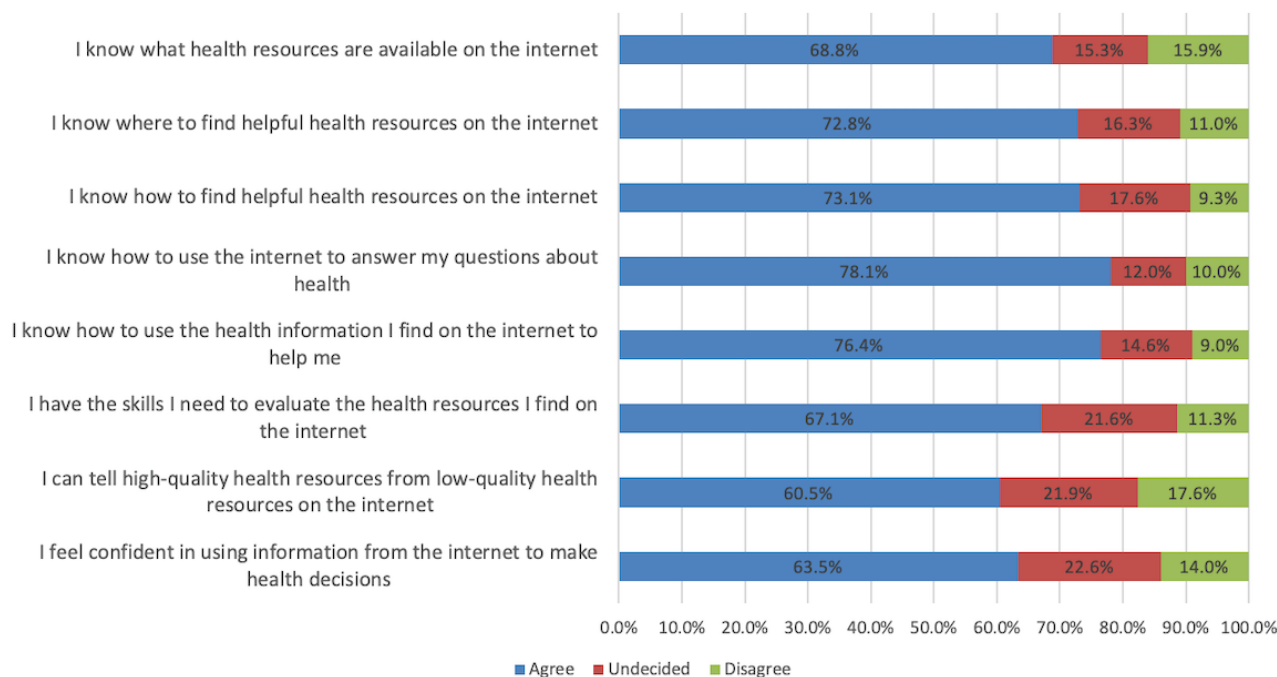
Mean eHEALS Scores by Demographic and Health Status Characteristics

Exploratory bivariate analysis showed that mean eHEALS scores were significantly lower in those aged ≥65 years (29.68, SD 6.55 vs 24.92, SD 8.40; difference –4.76, 95% CI –8.23 to –1.29; P=.01), and age was negatively correlated with eHEALS scores (Spearman correlation coefficient=–0.188; P=.001). Further, there were statistically significant higher mean eHEALS scores in those with college or university education versus high school education (30.45, SD 5.88 vs 25.27, SD 7.91; difference 5.19, 95% CI 2.58 to 7.79; P<.001), and those who preferred to receive written health information in English versus language other than English (30.38, SD 5.86 vs 22.22, SD 8.32; difference 8.17, 95% CI 5.45-10.8; P<.001; Table 1). There were also small but significantly higher scores in those free of hypertension or other chronic medical conditions.

Mean eHEALS Scores by Web-Based Health Information Seeking, Interest in eHealth Strategies, and Use of Health and Fitness Apps

In total, 74.4% (224/301) of participants used the internet several times per day, 71.4% (215/301) of participants sought web-based health information, and 31.6% (94/297) of smartphone or tablet owners used health and fitness apps. Of the internet users, 84.4% (254/301) were interested in future eHealth interventions. As shown in Table 2, the total eHEALS scores were significantly higher in those who used the internet several times per day; those using the web for health information; those using social media, Twitter, YouTube, and health and fitness apps; and those who showed interest in future eHealth interventions.

Table 2. Mean overall eHEALS[a] scores according to technology use.

	Values (n=301), n (%)	eHEALS score, mean (SD)	Difference in eHEALS score (95% CI)	P value[b]
Internet use frequency				<.001[c]
About once per day or less	74 (24.6)	26.49 (7.57)	Referent	
Several times per day	224 (74.4)	30.24 (6.17)	3.76 (1.83 to 5.68)	
Missing	3 (1)	N/A[d]	N/A	
Internet use for health information				.001[e]
No	84 (27.9)	26.37 (7.48)	Referent	
Yes	215 (71.4)	30.34 (6.23)	3.98 (2.16 to 5.78)	
Missing	2 (0.7)	N/A	N/A	
Social media use				.007[e]
No	68 (22.6)	26.97 (7.74)	Referent	
Yes	231 (76.7)	30.00 (6.25)	3.03 (1.06 to 5.06)	
Missing	2 (0.7)	N/A	N/A	
Twitter use				.03[e]
No	248 (82.4)	28.93 (6.66)	Referent	
Yes	50 (16.6)	31.48 (6.43)	2.55 (0.65 to 4.53)	
Missing	3 (1)	N/A	N/A	
YouTube				.047[e]
No	45 (15)	27.36 (7.56)	Referent	
Yes	253 (84.1)	29.70 (6.49)	2.34 (0.04 to 4.81)	
Missing	3 (1)	N/A	N/A	
Interested in future eHealth support[f]				.047[g]
No	47 (15.6)	26.91 (9.15)	Referent	
Yes	254 (84.4)	29.71 (6.25)	2.79 (0.3 to 5.57)	
Health and fitness apps[h]				.001[e]
No	199 (67)	27.10 (8.07)	Referent	
Yes	94 (31.6)	30.90 (6.90)	3.81 (2.07 to 5.73)	
Missing	4 (1.3)	N/A	N/A	

[a]eHEALS: eHealth Literacy Scale.

[b]P value is for comparison of eHEALS scores.

[c]Welch *F test* on trimmed mean and Winsorized variance as data violated constant variance assumption and were far from normality.

[d]N/A: not applicable.

[e]The bootstrap method with 1000 repetitions was used, as the data differed substantially from the normal distribution. For the bootstrap method, bias-corrected and accelerated CIs were reported.

[f]The 6 different modes of eHealth support in the future, included (1) accessing a webpage including a forum where you could connect with others like you; (2) accessing a YouTube channel for people with your conditions that has experts talking about best management; (3) using a smartphone app or wearable device that can monitor your condition, track your progress on your health goals, or provide reminders about when to take your medications; (4) following a specific Twitter account for your conditions; (5) signing up for personalized text messages providing health updates or reminders for your conditions; or (6) using a web-based education program.

[g]Welch *t* test.

[h]App use by 297 smartphone or tablet owners.

Characteristics Associated With eHealth Literacy: Quantile Regression

The results of the multivariable quantile regression using demographic, health status, and technology use variables for internet users are shown in Table 3. Language preference was the only variable independently associated with eHealth literacy. Expressing a preference for written health information in languages other than English reduced eHEALS scores by –6.0 (95% CI –9.61 to –2.39) points.

Table 3. Quantile regression of eHealth Literacy Scale scores through median for internet users (n=301).

	Internet users, coefficient (95% CI)
Age (years)	0.00 (–0.05 to 0.05)
Gender	
Male	Referent
Female	0.50 (–0.47 to 1.47)
Marital status	
Not married	Referent
Married	–0.50 (–2.04 to 1.04)
Education	
<High school	Referent
High School	7.69 (–13.17 to 28.56)
≥College	10.69 (–10.17 to 31.55)
Lived in Canada (years)	
>5	Referent
0-5	–1.0 (–2.09 to 0.09)
South Asian community	
Sikh	Referent
Hindu	0.00 (–1.03 to 1.03)
Other	–0.69 (–2.34 to 0.96)
Confidence in filling out medical forms	
>Not at all	Referent
Not at all	–4.81 (–20.19 to 10.57)
≥1 chronic condition	
No	Referent
Yes	–0.50 (–1.49 to 0.49)
Language preference	
English	Referent
Not English	–6.0 (–9.61 to –2.39)
Diabetes	
No	Referent
Yes	–0.50 (–2.69 to 1.69)
Amount of internet use	
Once per day or less	Referent
Several times per day	0.50 (–1.34 to 2.34)
Internet use for health information	
No	Referent
Yes	1.00 (–0.58 to 2.58)

Discussion

Principal Findings

Among our sample of primarily Sikh, highly educated internet users recruited from community pharmacies and family physician offices, the mean overall eHEALS score was 29.3 (SD 6.8) out of 40, suggesting high levels of eHealth literacy. Almost three quarters of the respondents agreed to at least 5 out of 8 eHEALS items and were categorized as having adequate eHealth literacy [49]. A large proportion of respondents (72%-78%) in this study reported knowing how to use the internet to answer health questions, how to use the health information found on the internet for support, and how to find helpful resources on the internet, whereas fewer (61%-67%) were confident in evaluating and using health information from the internet. Although there were statistically significant differences in eHEALS scores according to age, educational achievement, language preference, and the presence of hypertension or other chronic medical conditions, language preference was the only independent variable associated with eHealth literacy.

Comparison With Prior Work

Level of eHealth Literacy

Most comparable with our study, is a case study conducted by Zibrik et al [59], which used quantitative and qualitative methods to explore eHealth literacy in a convenience sample of 896 established Chinese and Punjabi immigrant seniors recruited from public health events in British Columbia, Canada. Although overall and item-specific eHEALS scores were not reported in their paper, the authors concluded that their sample had low eHealth literacy levels compounded by challenges related to language, culture, attitude, and accessibility. The discrepancy between this conclusion and our results can be attributed to the fact that we focused our analysis on internet users and excluded eHEALS scores from internet nonusers, whereas Zibrik included all participants regardless of internet use status. Another Canadian survey assessing digital health literacy in South Asian women suggested high digital health literacy via high rates of mobile phone internet use but was only reported in abstract form [60].

Our results compare favorably to existing studies that looked at measures of eHealth literacy among internet users in the general population, reporting mean overall eHEALS scores ranging between 24 and 30 out of 40 [26,49,61-63]. For example, in a 2014 survey, Milne et al [49] reported an overall eHEALS score of 24.0 (SD not reported) among 83 primary lung cancer survivors at a cancer center in Toronto, Ontario, 78% of whom had access to e-resources. Only 34% of lung cancer survivors in Milne's study agreed with 5 or more eHEALS items, whereas we found that 58% of all respondents agreed with 5 or more items [49]. More recently, James et al [63] reported a mean eHEALS score of 30.4 (SD 7.8) in a sample of 881 African American adults surveyed between April 2014 and January 2015 in North Central Florida.

In other studies focusing on eHealth literacy among internet users, as in our study, overall mean eHEALS scores ranged

from 16.1 (SD 4.25) to 30.34 (SD 5.30), depending on the population studied [16,22,50,64-68]. Our finding that the items with the lowest level of agreement among internet users are related to their ability to evaluate the quality of web-based health resources is highly consistent with the findings of several other studies [63,65,66,69].

Variables Associated With eHealth Literacy

Several studies have explored variables associated with eHealth literacy [16,17,22,34,50,62,65]. The most commonly reported are age [16,22,50,65], education [16,65], and markers of frequency or degree of internet use [17,22,34,62]. However, other associations, including the number of devices used to access web-based health information [16] and language [65], have also been reported. Conflicting results led Richtering et al [17] to conclude that variables associated with eHealth literacy are largely dependent on population. For example, in their study examining eHealth literacy in 453 participants enrolled in a randomized controlled trial of consumer-focused eHealth for cardiovascular risk management in primary care in Australia, they showed that the frequency of internet use was the sole predictor of eHealth literacy [17]. Focusing specifically on underserved populations, a systematic review of the research suggests that internet use experience, urban dwelling, higher income, overall health literacy, and higher education are associated with higher eHealth literacy [34]. Notably, Zibrik et al reported that college-level education and female gender were associated with higher eHEALS scores in 545 Punjabi seniors included in their Canadian study [59].

By contrast, we found that language preference was the only variable independently associated with eHealth literacy. This could be a proxy for acculturation, as years living in Canada were not an independent predictor in our model. Limited English proficiency is a widely cited barrier in studies examining health care and eHealth access in immigrant or minority populations [59]. However, contradictory evidence exists in this area, where language was not found to be an independent predictor of eHealth literacy in a Canadian study by Milne [49]. However, being a non-English speaker was significantly associated with lower eHealth literacy in an American study that included a significant number of Hispanic and African American parents whose children have special health care needs [65]. Further, our work is in contrast to that of Neter et al, which suggests that language difficulties should manifest as lower health literacy, rather than eHealth literacy, as health literacy as a concept is anchored in a cultural and language context, whereas eHealth literacy is anchored in the empowering and capital-enhancing qualities of the internet [37]. Our finding that educational status was not associated with eHealth literacy may be a result of the highly educated survey sample.

Implications for Practice and Research

First, although our results suggest that a large proportion of South Asian internet users have adequate eHealth literacy, there remains a sizable minority in our study group whose eHealth literacy can be strengthened. Understanding levels of eHealth literacy is even more important in 2022 than in 2014, as the COVID-19 pandemic has globally increased the remote internet-based delivery of health care [19]. Second, our results

suggest that there is an opportunity to improve internet users' ability to differentiate high- from low-quality health information. Beyond assessing web-based health information about diabetes, cardiovascular disease, physical activity, and diet or nutrition, the COVID-19 infodemic is another example that illustrates the need for individuals to be better able to identify medical misinformation on the internet [21]. Training individuals to recognize misinformation could be one way to increase their eHealth literacy. Another approach may include health care organizations and provider support for patients to navigate and access high-quality and reliable web-based health information and resources via social media [70]. Finally, given that language preference was the only predictor of eHealth literacy, and this is likely a nonmodifiable factor, we suggest that interventions to improve eHealth literacy should be targeted toward those who have low English language proficiency and delivered in the individual's preferred language. Strategies involving peers, friends, and family members may also be effective.

Our study suggests several areas for future research. First, further work with representative samples should compare levels of health and eHealth literacy using newly available tools among recent and established Canadian immigrants and nonimmigrants similar to that of Neter et al [37]. Second, work should be done to develop theory driven, cultural, and language-tailored interventions to increase the uptake of eHealth interventions and improve eHealth literacy. Finally, high-quality randomized controlled trials are necessary to evaluate theory-based eHealth literacy interventions in ethnocultural minority and immigrant populations, and their impact on health outcomes.

Strengths and Limitations

At the time the survey was conducted in 2014, our work was unique, and to our knowledge, there are only 2 other published studies exploring eHealth literacy in members of the South Asian community in Canada [59,60]. Despite this, our study has several limitations. First, our data were collected in 2014 and are not likely to reflect the current use of digital health technologies or eHealth literacy. Second, as nonprobability sampling was used and only a subset of the entire survey sample completed the eHEALS assessment, we were unable to generalize our results to the larger South Asian population, as the sample is not representative. Furthermore, our results primarily pertain to the English- and Punjabi-speaking Sikh community, as we did not translate our survey into other commonly spoken languages (eg, Hindi and Urdu). Third, we did not formally validate the Punjabi version of the eHEALS, and there may be issues with conceptual translation and some variability in administration. However, few eHealth literacy assessments have been completed in Punjabi, and we used established procedures for translating survey items. We reported high internal consistency for the 8 items in this survey, and we had 2 trained research assistants who administered over 65% of the surveys. Fourth, eHEALS measures perceived eHealth literacy rather than actual eHealth literacy or skills as measured by performance tests [67,71]. In addition, it is worth noting that the tool does not address the ability to use Web 2.0 functionalities such as social media, mobile devices, and health and fitness apps for health behavior change purposes as do new tools [72]. Fifth, as our survey was primarily administered in person, social desirability bias may overinflate reported eHEALS scores and estimates of device ownership, internet use, and willingness to use future eHealth tools, whereas self-reporting may introduce recall bias in outcome and demographic variables. Finally, we recognize that the question relating to language preference for written health information could have been improved by asking about the primary language spoken in the home and that the use of a single health literacy screening question rather than a full health literacy questionnaire is not optimal.

Conclusions

This cross-sectional study in a subset of e-Patient Project survey respondents provides insights for clinicians and researchers on the levels and variables associated with eHealth literacy in a sample of South Asian adults living in a major Canadian city. Preferring written health information in languages other than English was the only independent variable associated with reduced eHealth literacy in our sample. Our results suggest that respondents may still benefit from interventions targeting skills to evaluate web-based health resources and that linguistically and culturally appropriate interventions are required to improve eHealth literacy.

XSL·FO

RenderX

Authors' Contributions

MJM, SD, and CAJ contributed to conception and design. MJM prepared the first draft of the manuscript and contributed to acquisition of data, analysis, and interpretation. MJM, SD, and CAJ were involved in critical revision of the manuscript for intellectual content. MJM, SD, and CAJ handled the final approval of the version to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1
English version of the e-Patient Project Survey and eHealth Literacy Scale.
[DOCX File , 49 KB - formative_v6i3e29955_app1.docx]

Multimedia Appendix 2
Punjabi translation of the eHealth Literacy Scale.
[PDF File (Adobe PDF File), 53 KB - formative_v6i3e29955_app2.pdf]

Multimedia Appendix 3
eHealth Literacy Scale mean item scores, scale reliability, and principal component analysis (n=301).
[DOCX File , 15 KB - formative_v6i3e29955_app3.docx]

Multimedia Appendix 4
Characteristics of internet nonuser eHealth Literacy Scale completers (n=83).
[DOCX File , 22 KB - formative_v6i3e29955_app4.docx]

References

1. Lupton D. The digitally engaged patient: self-monitoring and self-care in the digital health era. Soc Theory Health 2013 Jun 19;11(3):256-270. [doi: 10.1057/sth.2013.10]
2. Canada Health Infoway. URL: https://www.infoway-inforoute.ca/en/component/edocman/resources/technical-documents/architecture/2944-digital-health-blueprint-enabling-coordinated-collaborative-health-care [accessed 2022-03-09]
3. Meier CA, Fitzgerald MC, Smith JM. eHealth: extending, enhancing, and evolving health care. Annu Rev Biomed Eng 2013;15:359-382. [doi: 10.1146/annurev-bioeng-071812-152350] [Medline: 23683088]
4. Oldenburg B, Taylor CB, O'Neil A, Cocker F, Cameron LD. Using new technologies to improve the prevention and management of chronic conditions in populations. Annu Rev Public Health 2015 Mar 18;36:483-505. [doi: 10.1146/annurev-publhealth-031914-122848] [Medline: 25581147]
5. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013 Dec;10(1):e1001362-e1001364 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
6. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
7. Bacigalupo R, Cudd P, Littlewood C, Bissell P, Hawley MS, Buckley Woods H. Interventions employing mobile technology for overweight and obesity: an early systematic review of randomized controlled trials. Obes Rev 2013 Apr;14(4):279-291 [FREE Full text] [doi: 10.1111/obr.12006] [Medline: 23167478]
8. Wieland LS, Falzon L, Sciamanna CN, Trudeau KJ, Brodney S, Schwartz JE, et al. Interactive computer-based interventions for weight loss or weight maintenance in overweight or obese people. Cochrane Database Syst Rev 2012 Aug 15(8):CD007675 [FREE Full text] [doi: 10.1002/14651858.CD007675.pub2] [Medline: 22895964]
9. Chang T, Chopra V, Zhang C, Woolford SJ. The role of social media in online weight management: systematic review. J Med Internet Res 2013 Nov 28;15(11):e262 [FREE Full text] [doi: 10.2196/jmir.2852] [Medline: 24287455]
10. Fanning J, Mullen SP, McAuley E. Increasing physical activity with mobile devices: a meta-analysis. J Med Internet Res 2012 Nov 21;14(6):e161 [FREE Full text] [doi: 10.2196/jmir.2171] [Medline: 23171838]
11. Foster C, Richards J, Thorogood M, Hillsdon M. Remote and web 2.0 interventions for promoting physical activity. Cochrane Database Syst Rev 2013 Sep 30;9:CD010395. [doi: 10.1002/14651858.CD010395.pub2] [Medline: 24085594]
12. Holtz B, Lauckner C. Diabetes management via mobile phones: a systematic review. Telemed J E Health 2012 Apr;18(3):175-184. [doi: 10.1089/tmj.2011.0119] [Medline: 22356525]

XSL•FO
RenderX

13. Saffari M, Ghanizadeh G, Koenig HG. Health education via mobile text messaging for glycemic control in adults with type 2 diabetes: a systematic review and meta-analysis. Prim Care Diabetes 2014 Dec;8(4):275-285. [doi: 10.1016/j.pcd.2014.03.004] [Medline: 24793589]

14. Pal K, Eastwood SV, Michie S, Farmer AJ, Barnard ML, Peacock R, et al. Computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus. Cochrane Database Syst Rev 2013;3:CD008776. [doi: 10.1002/14651858.CD008776.pub2] [Medline: 23543567]

15. Watkins I, Xie B. eHealth literacy interventions for older adults: a systematic review of the literature. J Med Internet Res 2014;16(11):e225 [FREE Full text] [doi: 10.2196/jmir.3318] [Medline: 25386719]

16. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]

17. Richtering SS, Hyun K, Neubeck L, Coorey G, Chalmers J, Usherwood T, et al. eHealth literacy: predictors in a population with moderate-to-high cardiovascular risk. JMIR Hum Factors 2017 Jan 27;4(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.6217] [Medline: 28130203]

18. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]

19. Glazier RH, Green ME, Wu FC, Frymire E, Kopp A, Kiran T. Shifts in office and virtual primary care during the early COVID-19 pandemic in Ontario, Canada. CMAJ 2021 Feb 08;193(6):E200-E210 [FREE Full text] [doi: 10.1503/cmaj.202303] [Medline: 33558406]

20. Brørs G, Norman CD, Norekvål TM. Accelerated importance of eHealth literacy in the COVID-19 outbreak and beyond. Eur J Cardiovasc Nurs 2020 Aug;19(6):458-461 [FREE Full text] [doi: 10.1177/1474515120941307] [Medline: 32667217]

21. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, et al. The COVID-19 social media infodemic. Sci Rep 2020 Oct 06;10(1):16598 [FREE Full text] [doi: 10.1038/s41598-020-73510-5] [Medline: 33024152]

22. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: internet use patterns, eHealth literacy, and attitudes toward computer/internet use. J Med Internet Res 2013 May;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]

23. Czaja SJ, Sharit J, Lee CC, Nair SN, Hernández MA, Arana N, et al. Factors influencing use of an e-health website in a community sample of older adults. J Am Med Inform Assoc 2013;20(2):277-284 [FREE Full text] [doi: 10.1136/amiajnl-2012-000876] [Medline: 22802269]

24. Jensen JD, King AJ, Davis LA, Guntzviller LM. Utilization of internet technology by low-income adults: the role of health literacy, health numeracy, and computer assistance. J Aging Health 2010 Sep;22(6):804-826. [doi: 10.1177/0898264310366161] [Medline: 20495159]

25. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]

26. Robinson C, Graham J. Perceived Internet health literacy of HIV-positive people through the provision of a computer and internet health education intervention. Health Info Libr J 2010 Dec;27(4):295-303 [FREE Full text] [doi: 10.1111/j.1471-1842.2010.00898.x] [Medline: 21050372]

27. Xie B. Effects of an eHealth literacy intervention for older adults. J Med Internet Res 2011 Nov 03;13(4):e90 [FREE Full text] [doi: 10.2196/jmir.1880] [Medline: 22052161]

28. Census of population. Statistics Canada. URL: http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/demo50a-eng.htm [accessed 2022-03-09]

29. Fernando E, Razak F, Lear SA, Anand SS. Cardiovascular disease in South Asian migrants. Can J Cardiol 2015 Sep;31(9):1139-1150. [doi: 10.1016/j.cjca.2015.06.008] [Medline: 26321436]

30. Shah BR, Cauch-Dudek K, Anand SS, Austin PC, Manuel DG, Hux JE. Absence of disparities in the quality of primary diabetes care for South Asians and Chinese in an urban Canadian setting. Diabetes Care 2012 Apr;35(4):794-796 [FREE Full text] [doi: 10.2337/dc11-1845] [Medline: 22323411]

31. Gray J, Millett C, Saxena S, Netuveli G, Khunti K, Majeed A. Ethnicity and quality of diabetes care in a health system with universal coverage: population-based cross-sectional survey in primary care. J Gen Intern Med 2007 Sep;22(9):1317-1320 [FREE Full text] [doi: 10.1007/s11606-007-0267-4] [Medline: 17594128]

32. Sohal T, Sohal P, King-Shier KM, Khan NA. Barriers and facilitators for type-2 diabetes management in South Asians: a systematic review. PLoS One 2015;10(9):e0136202 [FREE Full text] [doi: 10.1371/journal.pone.0136202] [Medline: 26383535]

33. Shariff AI, Kumar N, Yancy WS, Corsino L. Type 2 diabetes and atherosclerotic cardiovascular disease in South Asians: a unique population with a growing challenge. Curr Diab Rep 2020 Jan 30;20(1):4. [doi: 10.1007/s11892-020-1291-6] [Medline: 32002674]

34. Chesser A, Burke A, Reyes J, Rohrberg T. Navigating the digital divide: a systematic review of eHealth literacy in underserved populations in the United States. Inform Health Soc Care 2016;41(1):1-19. [doi: 10.3109/17538157.2014.948171] [Medline: 25710808]

35.   Ng E, Omariba D. Immigration, generational status and health literacy in Canada. Health Educ J 2013 Nov 27;73(6):668-682. [doi: 10.1177/0017896913511809]

36.   Villadsen SF, Hadi H, Ismail I, Osborne RH, Ekstrøm CT, Kayser L. ehealth literacy and health literacy among immigrants and their descendants compared with women of Danish origin: a cross-sectional study using a multidimensional approach among pregnant women. BMJ Open 2020 May 07;10(5):e037076 [FREE Full text] [doi: 10.1136/bmjopen-2020-037076] [Medline: 32385065]

37.   Neter E, Brainin E, Baron-Epel O. Group differences in health literacy are ameliorated in ehealth literacy. Health Psychol Behav Med 2021 May 15;9(1):480-497 [FREE Full text] [doi: 10.1080/21642850.2021.1926256] [Medline: 34104571]

38.   Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]

39.   Yoon H, Jang Y, Vaughan PW, Garcia M. Older adults' internet use for health information: digital divide by race/ethnicity and socioeconomic status. J Appl Gerontol 2020 Jan;39(1):105-110. [doi: 10.1177/0733464818770772] [Medline: 29661052]

40.   Makowsky MJ, Jones CA, Davachi S. Prevalence and predictors of health-related internet and digital device use in a sample of South Asian adults in Edmonton, Alberta, Canada: results from a 2014 community-based survey. JMIR Public Health Surveill 2021 Jan 08;7(1):e20671 [FREE Full text] [doi: 10.2196/20671] [Medline: 33416506]

41.   Mobile health 2012. Pew Research Center. URL: http://www.pewinternet.org/2012/11/08/mobile-health-2012/ [accessed 2022-03-09]

42.   Health online 2013. Pew Research Center. URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2022-03-09]

43.   Tracking for health. Pew Research Center. URL: http://www.pewinternet.org/2013/01/28/tracking-for-health/ [accessed 2022-03-09]

44.   Canadian internet use survey, 2012. Statistics Canada. URL: http://www.statcan.gc.ca/daily-quotidien/131126/dq131126d-eng.pdf [accessed 2022-03-09]

45.   Collins SA, Currie LM, Bakken S, Vawdrey DK, Stone PW. Health literacy screening instruments for eHealth applications: a systematic review. J Biomed Inform 2012 Jun;45(3):598-607 [FREE Full text] [doi: 10.1016/j.jbi.2012.04.001] [Medline: 22521719]

46.   Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]

47.   Process of translation and adaptation of instruments. World Health Organization. URL: http://www.who.int/substance_abuse/research_tools/translation/en/ [accessed 2022-03-09]

48.   Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med 2004 Sep;36(8):588-594 [FREE Full text] [Medline: 15343421]

49.   Milne RA, Puts MT, Papadakos J, Le LW, Milne VC, Hope AJ, et al. Predictors of high eHealth literacy in primary lung cancer survivors. J Cancer Educ 2015 Dec;30(4):685-692. [doi: 10.1007/s13187-014-0744-5] [Medline: 25355524]

50.   Cherid C, Baghdadli A, Wall M, Mayo NE, Berry G, Harvey EJ, et al. Current level of technology use, health and eHealth literacy in older Canadians with a recent fracture-a survey in orthopedic clinics. Osteoporos Int 2020 Jul 28;31(7):1333-1340. [doi: 10.1007/s00198-020-05359-3] [Medline: 32112118]

51.   Moser BK, Stevens GR. Homogeneity of variance in the two-sample means test. Am Statistician 1992 Feb;46(1):19. [doi: 10.2307/2684403]

52.   Ruxton GD. The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. Behav Ecol 2006;17(4):688-690 [FREE Full text]

53.   Fagerland MW, Sandvik L. The Wilcoxon-Mann-Whitney test under scrutiny. Stat Med 2009 May 01;28(10):1487-1497. [doi: 10.1002/sim.3561] [Medline: 19247980]

54.   Lydersen S. Statistical review: frequently given comments. Ann Rheum Dis 2015 Feb 26;74(2):323-325. [doi: 10.1136/annrheumdis-2014-206186] [Medline: 25261576]

55.   Wilcox RR. ANOVA - the practical important of heteroscedastic methods, using trimmed means versus means, and designing simulation studies. British J Mathematical Stat Psychol 1995;48(1):99-114. [doi: 10.1111/j.2044-8317.1995.tb01052.x]

56.   Wilcox RR. Introduction to Robust Estimation and Hypothesis Testing. 3rd Edition. Amsterdam The Netherlands: Academic Press, Elsevier Inc; 2012.

57.   Cribbie RA, Fiksenbaum L, Keselman HJ, Wilcox RR. Effect of non-normality on test statistics for one-way independent groups designs. Br J Math Stat Psychol 2012 Feb;65(1):56-73. [doi: 10.1111/j.2044-8317.2011.02014.x] [Medline: 21426313]

58.   Koenker R. Quantile Regression. Cambridge; New York: Cambridge University Press; 2005.

59.   Zibrik L, Khan S, Bangar N, Stacy E, Novak Lauscher H, Ho K. Patient and community centered eHealth: exploring eHealth barriers and facilitators for chronic disease self-management within British Columbia's immigrant Chinese and Punjabi seniors. Health Policy Technol 2015 Dec;4(4):348-356. [doi: 10.1016/j.hlpt.2015.08.002]

60.   Zhang S, Pedersen H, Smith L, Sarai RC, Niekerk D, Lee M. P830 - Feasibility of an online HPV self-collection screening program in Canada: digital health literacy in south asian women. Sexually Transmitted Infections 2019;95(Suppl 1):A349. [doi: 10.1136/sextrans-2019-sti.875]

61. Xesfingi S, Vozikis A. eHealth literacy: in the quest of the contributing factors. Interact J Med Res 2016 May 25;5(2):e16 [FREE Full text] [doi: 10.2196/ijmr.4749] [Medline: 27226146]

62. Wong DK, Cheung M. Online health information seeking and eHealth literacy among patients attending a primary care clinic in Hong Kong: a cross-sectional survey. J Med Internet Res 2019 Mar 27;21(3):e10831 [FREE Full text] [doi: 10.2196/10831] [Medline: 30916666]

63. James DC, Harville C. Ehealth literacy, online help-seeking behavior, and willingness to participate in mHealth chronic disease research among African Americans, Florida, 2014-2015. Prev Chronic Dis 2016 Nov 17;13:E156 [FREE Full text] [doi: 10.5888/pcd13.160210] [Medline: 27854421]

64. Kim S, Son Y. Relationships between eHealth literacy and health behaviors in Korean adults. Comput Inform Nurs 2017 Feb;35(2):84-90. [doi: 10.1097/CIN.0000000000000255] [Medline: 27258808]

65. Knapp C, Madden V, Wang H, Sloyer P, Shenkman E. Internet use and eHealth literacy of low-income parents whose children have special health care needs. J Med Internet Res 2011 Sep 29;13(3):e75 [FREE Full text] [doi: 10.2196/jmir.1697] [Medline: 21960017]

66. Park H, Cormier E, Gordon G, Baeg JH. Identifying health consumers' eHealth literacy to decrease disparities in accessing eHealth information. Comput Inform Nurs 2016 Feb;34(2):71-6; quiz 99. [doi: 10.1097/CIN.0000000000000205] [Medline: 26657619]

67. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 09;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]

68. Paige SR, Krieger JL, Stellefson M, Alber JM. eHealth literacy in chronic disease patients: an item response theory analysis of the eHealth literacy scale (eHEALS). Patient Educ Couns 2017 Feb;100(2):320-326 [FREE Full text] [doi: 10.1016/j.pec.2016.09.008] [Medline: 27658660]

69. Manafò E, Wong S. Assessing the eHealth literacy skills of older adults: a preliminary study. J Consum Health Internet 2012 Oct;16(4):369-381. [doi: 10.1080/15398285.2012.701163]

70. Together against misinformation. Science Up First. URL: https://www.scienceupfirst.com [accessed 2022-03-09]

71. van der Vaart R, Drossaert CH, de Heus M, Taal E, van de Laar MA. Measuring actual eHealth literacy among patients with rheumatic diseases: a qualitative analysis of problems encountered using Health 1.0 and Health 2.0 applications. J Med Internet Res 2013 Feb 11;15(2):e27 [FREE Full text] [doi: 10.2196/jmir.2428] [Medline: 23399720]

72. van der Vaart R, Drossaert C. Development of the digital health literacy instrument: measuring a broad spectrum of health 1.0 and health 2.0 skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]

Abbreviations

eHEALS: eHealth Literacy Scale
mHealth: mobile health

<u>Original Paper</u>

Patient Perspectives of Inpatient Telemedicine During the COVID-19 Pandemic: Qualitative Assessment

Stacie Vilendrer[1], MD, MBA, MS; Sarah Sackeyfio[2]; Eliel Akinbami[3], PhD; Roy Ghosh[3]; Jacklyn Ha Luu[4]; Divya Pathak[5]; Masahiro Shimada[6]; Emmanuelle Elise Williamson[7]; Lisa Shieh[8], MD, PhD

[1]Division of Primary Care and Population Health, Stanford University School of Medicine, Stanford, CA, United States

[2]Department of Physics, Stanford University, Stanford, CA, United States

[3]Department of Bioengineering, Stanford University, Stanford, CA, United States

[4]Department of Bioinformatics, Stanford University School of Medicine, Stanford, CA, United States

[5]Department of Biochemistry, Stanford University, Stanford, CA, United States

[6]Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA, United States

[7]Department of Engineering, Stanford University, Stanford, CA, United States

[8]Division of Hospital Medicine, Stanford University School of Medicine, Stanford, CA, United States

Corresponding Author:
Stacie Vilendrer, MD, MBA, MS
Division of Primary Care and Population Health
Stanford University School of Medicine
1265 Welch Rd
Stanford, CA, 94305
United States
Phone: 1 952 913 8883
Email: staciev@stanford.edu

Abstract

Background: Telemedicine has been adopted in the inpatient setting to facilitate clinical interactions between on-site clinicians and isolated hospitalized patients. Such remote interactions have the potential to reduce pathogen exposure and use of personal protective equipment but may also pose new safety concerns given prior evidence that isolated patients can receive suboptimal care. Formal evaluations of the use and practical acceptance of inpatient telemedicine among hospitalized patients are lacking.

Objective: We aimed to evaluate the experience of patients hospitalized for COVID-19 with inpatient telemedicine introduced as an infection control measure during the pandemic.

Methods: We conducted a qualitative evaluation in a COVID-19 designated non–intensive care hospital unit at a large academic health center (Stanford Health Care) from October 2020 through January 2021. Semistructured qualitative interviews focused on patient experience, impact on quality of care, communication, and mental health. Purposive sampling was used to recruit participants representing diversity across varying demographics until thematic saturation was reached. Interview transcripts were qualitatively analyzed using an inductive-deductive approach.

Results: Interviews with 20 hospitalized patients suggested that nonemergency clinical care and bridging to in-person care comprised the majority of inpatient telemedicine use. Nurses were reported to enter the room and call on the tablet far more frequently than physicians, who typically entered the room at least daily. Patients reported broad acceptance of the technology, citing improved convenience and reduced anxiety, but preferred in-person care where possible. Quality of care was believed to be similar to in-person care with the exception of a few patients who wanted more frequent in-person examinations. Ongoing challenges included low audio volume, shifting tablet location, and inconsistent verbal introductions from the clinical team.

Conclusions: Patient experiences with inpatient telemedicine were largely favorable. Although most patients expressed a preference for in-person care, telemedicine was acceptable given the circumstances associated with the COVID-19 pandemic. Improvements in technical and care team use may enhance acceptability. Further evaluation is needed to understand the impact of inpatient telemedicine and the optimal balance between in-person and virtual care in the hospital setting.

Introduction

The COVID-19 pandemic created an unprecedented challenge for health systems to provide high-quality care to potentially contagious patients while simultaneously keeping their workforce and uninfected patients safe [1]. In response, telemedicine was widely adopted worldwide in outpatient settings [2]. Some further adopted "inpatient telemedicine" in acute care settings to facilitate clinical interactions between on-site clinicians and isolated inpatients for the purpose of infection control and to reduce the use of personal protective equipment (PPE) [3-7]. Early evidence points to the feasibility of inpatient telemedicine as an infection control measure [3-5], but formal evaluations into their use and practical acceptance among stakeholders, particularly patients, are limited.

In the outpatient setting, robust literature suggests overall patient satisfaction with telemedicine, including in general medicine [8], urgent care [9], and specialized settings [10-12]. Factors driving increased patient satisfaction include improved outcomes, decreased travel time, and improved communication [13]. Adequate training of the staff providing telemedicine services can also improve patient satisfaction [14]. Patient acceptance of telemedicine has persisted through the pandemic [15-18], with the highest levels of adoption observed among young patients [19]. However, high rates of patient satisfaction with telemedicine in the outpatient setting do not necessarily translate to the inpatient setting, given significant differences in the clinical environment, acuity of disease, frequency of clinical interactions, and the multidisciplinary team required to care for the patient.

Prior to the COVID-19 pandemic, the use of telemedicine in inpatient settings was limited to connecting rural areas with remote expertise [20,21], particularly specialty care [14,22-24]. The COVID-19 pandemic removed traditional barriers to telemedicine adoption, including staff resistance to change, and lack of reimbursement [13]. Since the pandemic, hospitalists [3,5], intensivists [4,6], and specialists [25,26] have used inpatient telemedicine to continue their clinical duties while minimizing pathogen spread. This novel use of inpatient telemedicine is still being explored: in one academic medical setting, the technology was broadly accepted by clinicians and staff but required nurses to drastically alter their workflow practices [27]. In another setting, in-room video technology saw the greatest adoption among patients who used it to communicate with their loved ones [28]. Little else is formally known about how patients perceive inpatient telemedicine, particularly under isolation conditions that a COVID-19 diagnosis necessitates.

This is unfortunate, as the patient perspective is integral to ensuring robust patient–provider communication, the lack of which can propagate diagnostic errors and adverse patient outcomes [29-31]. Isolated patients may be more likely to receive substandard care, particularly related to patient safety [32] and staff responsiveness to time-sensitive needs [33,34]. Further, patients hospitalized under isolation precautions face unique challenges, including negative feelings associated with loneliness, stigma, and fear [31-37]. Incorporating patient insights into the design of technological solutions within the patient environment is needed to support patient autonomy, human connection, and a sense familiarity that is otherwise lost during isolation [36-38].

We therefore aim to understand patient experience as it relates to clinical use and patient acceptance of inpatient telemedicine for infection control in order to optimize technical setup and clinical workflows.

Methods

Methods Overview

We conducted a qualitative evaluation of inpatient telemedicine use on a COVID-19 designated inpatient unit at a large academic medical center. We analyzed qualitative interviews with stable patients hospitalized for COVID-19 to understand the implementation outcomes related to the practical use and acceptance of the inpatient telemedicine solution from the patients' perspective.

Setting and Population

In March 2020, a large academic health center (Stanford Health Care) designated an inpatient non–intensive care unit to care for admitted patients with confirmed or suspected COVID-19. Each patient received a tablet (iPad; Apple Inc) set up to automatically receive web conference calls (Zoom Video Communications, Inc) from one of two desktops in private workrooms on the unit, all with dual audio and video functionality. Other setting details have been previously described [3] and can be found in Multimedia Appendix 1.

The target population included stable patients who were admitted to the non–intensive care COVID-19 unit. Purposive sampling methods were used to recruit participants representing diversity across age, sex, race and ethnicity, and language until thematic saturation was reached [39].

Data Collection and Analysis

The semistructured patient interview protocol, analysis, and presentation of findings were informed through deductive themes drawn from prior qualitative work [27] and the Implementation Outcomes Framework (Multimedia Appendix 1) [40]. Interview transcripts were reviewed independently by the lead author (SV), and key themes and inductive coding that emerged were then validated using a consensus approach that included the analytic (DP, MS, JHL, and RG) and interview (SS, EA, and EEW) teams.

Ethical Considerations and Informed Consent

The Stanford Institutional Review Board determined that this project did not qualify as human subjects research (protocol

55927). All participants provided verbal consent to proceed and were informed their participation was voluntary and confidential.

Results

Participant Characteristics

A total of 20 interviews were conducted with inpatients undergoing treatment for COVID-19 without intensive care needs. Participants tended to be male (11/20, 55%); aged 50-69 years (15/20, 75%); Latino, White, or Asian, Indian, or Pacific Islander (6/20, 30% each), and spoke primarily English (16/20, 80%) (Multimedia Appendix 1).

Reported Use and Acceptance of Inpatient Telemedicine From the Patients' Perspective

Telemedicine use varied on the basis of the clinical context and the type of clinician providing care, though it was broadly accepted by patients given the COVID-19–related isolation precautions (Table 1). Predominant use included nonemergency clinical care and bridging to in-person care—such as when a patient triggered the bedside alert button. In this setting, one nurse was reported to initiate a telemedicine encounter to visually connect with the patient while another donned PPE to evaluate the patient in person.

Typical frequency of virtual encounters ranged from 1 to 3 times a day, though one patient reported being contacted up to 10 times per day via the tablet device and 10 times per day in person. Nurses were reported to both enter the room and call on the tablet more frequently than physicians. However, most patients reported at least one in-person encounter daily from their doctor, with a few reporting multiple in-person encounters when their needs were greater. A few patients reported their physician primarily communicated via the tablet device, and a minority perceived there were some days on which they did not speak with their doctor either in person or via the tablet device.

Patients generally accepted the use of inpatient telemedicine given their circumstances requiring isolation precautions. A desire to prevent infection transmission was recognized by several patients: "It's helpful that I can talk to multiple doctors without having to expose them to COVID" [Patient 9]. Some felt that telemedicine offered superior convenience, and the visual component was seen as adding value beyond an audio-only telephone encounter. The virtual interactions provided reassurance, which was reported to positively impact patient mood and mental health: "I love to hear from my doctors, it brought such great comfort for them to be updating me" [Patient 6].

Patients broadly reported that the quality of care they received with the integration of telemedicine into their hospitalization was similar to what they might have received without it.

However, some were in favor of an in-person encounter in accordance with the concern that some aspects of their care could be missed, particularly related to the in-person examination and human connection. Some felt that telemedicine changed the way they communicated with their care team: "In person is better…I can have a more in-depth dialogue, ask questions…When they're physically here, they can be there for me" [Patient 18]. Older patients, in particular, seemed to prefer in-person encounters. Only a handful of patients felt telemedicine did not compromise their connection with their care team.

Table 1. Patient perspectives on the use and acceptance of inpatient telemedicine in the context of the COVID-19 pandemic.

Learnings	Example quotations
Reported use of inpatient telemedicine	
Used primarily for nonemergency care and as a bridge to in-person care	• "We run the visits based on my bladder which is about every hour and a half. They [nurses] bring breakfast and then I pee and then they do my meds and stuff. Normally they don't come in first, [instead] they check on me on [the tablet] and then come in." [Patient 20]
Nurses reported to both enter the room and call on the tablet device	• "Five or six times a day [in-person nurse visits] and they [nurses] also call on the iPad…" [Patient 9]
Physicians sometimes used telemedicine to replace in-person encounters	• "At least once maybe twice a day [in-person physician encounter]. There were some days though where I didn't have any doctors come in and talk to me. They were just conversing with the staff regarding my wellbeing." [Patient 6]
Acceptance of inpatient telemedicine	
Generally accepted given the circumstances surrounding the COVID-19 pandemic	• "Of course there's nothing more effective than the person to person contact, but this is the safest, and the tech is so advanced it's like you're right in front of me" [Patient 15]
Telemedicine seen as improving convenience	• "I think it's great because we have more contact with [the] medical providers than any other time. You normally have to wait till they have time to come and talk to you…[but this way] seems to be a lot more efficient way to handle things" [Patient 2]
Immediate accessibility to clinical team via telemedicine offered reassurance	• "On the first day I went to sleep, woke up with anxiety real bad, feeling as though I was on my last breath. I pushed the [bedside alert] button and the nurses came on and then after that the [tablet] did come on they were looking at me, talking to me, and helped me calm down right away. They were like, 'Just breathe, someone's coming right in'." [Patient 16]
Visual component seen as adding value over a telephone call	• "…they [clinicians] are right there when I need them and I can physically see them versus waiting to talk to them via the phone." [Patient 8]
No perceived impact on quality of care overall with a minority concerned about the reduction in in-person examinations	• "Honestly I don't think it has changed the care because the nurse is always reporting to the doctor and then the doctor looks at all my labs and things so they are pretty on top of it and then they call me to let me know the plan…But I wouldn't say it's better it's like the same." [Patient 1] • "Most of time when we see doctors normally, they use the stethoscope, to listen to lungs, heart…That's the normal way to see a patient, so I just wonder, I'm completely okay with the [tablet], but somehow the doctor can't see [me] in person…the nurse never uses the [stethoscope] to listen to my lungs/heart. So that is something I am not comfortable with." [Patient 4]
Some reported loss of human connection	• "…on the internet you can say whatever you want; telemedicine is close to internet separation…the non-word ways of communication are much different than in person, tone, intonation…like I'm here but I don't see your whole body and vice versa. [It] changes [one's] mood..." [Patient 3]

Technical Factors Impacting Patient Acceptance

In general, the technology was reportedly "very simple to use" [Patient 5]. However, certain technical setup considerations impacted patient acceptability, including audio volume, tablet position both within the room and in relation to the patient, and the automatic turn-on feature. Several patients remarked that despite turning the tablet volume up to its maximum capacity, they still had difficulty hearing the clinical team, particularly if someone was speaking further away from the source microphone.

Furthermore, the tablet location and angle relative to the patient sometimes posed challenges. The tablet device was reportedly moved frequently to allow space for nurses doing their clinical work. This lack of a stable location worsened the experience for some patients:

> Mostly I don't like it. I don't understand them sometimes. Sometimes they move it towards my feet. It's too far from me, it should be near my face... [Patient 19]

There was not a clear consensus on where the tablet should be placed within the room, as another patient preferred it near the feet to optimize the visual component of the encounter.

Finally, patients were asked to comment on the automatic turn-on feature of the tablet device, which did not allow them to screen calls. Nearly all were comfortable with this feature, likening it to standard hospital care in which a clinician would simply walk into the room, only sometimes following a quick

knock. The one patient in his 20s, who expressed some concern over privacy, devised his own solution to turn the camera toward the wall temporarily while he showered.

Care Team Use Factors Impacting Patient Acceptance

The clinical team's use of the technology also impacted patient acceptance, specifically around patient orientation to technology, clinician etiquette regarding encounter introductions and use, and visual connection with each speaker. Patients described minimal or no orientation to the tablet. "They never told me about the iPad; I noticed when they came on-- 'Hello hello hello'" [Patient 10]. Where an orientation was reported to take place, it was minimal: "They pointed it out to me and said…this is the [tablet] for nurses and doctors to call you" [Patient 11]. While this brief orientation would have been preferred, it did not bother most patients. A preference for an orientation was more prevalent among patients who did not have prior familiarity with using tablet devices or web conferencing technology.

Opportunities were also noted for improved etiquette around introducing each member of the team and positioning the tablet toward the speaker:

> *I have a bunch of different doctors on my care team…[it is] much harder to remember each one, [as I] don't have visual clues. I see [them] only as 2D and honestly [there is an] unwillingness for everyone to get in front of the screen…one [clinician] is on the side, I hear their comments but I don't know who that is...* [Patient 3]

This concern was not prevalent among all patients but may have improved acceptability among the few patients who voiced it.

Discussion

Principal Findings

Patients' experience with telemedicine during an inpatient admission for COVID-19 was largely favorable. While most patients expressed a preference for in-person care where possible, telemedicine was an acceptable alternative given COVID-19 isolation precautions. Both nurses and physicians regularly used the technology to communicate with patients, thus validating past work [27] in which the balance of in-person versus telemedicine workflows was guided by acuity of clinical need. Telemedicine appeared to serve as a supplementary point of contact between in-person encounters, which patients primarily saw as a benefit to their overall experience and mental health. The technology reportedly played a meaningful role in mitigating the fear and anxiety associated with isolation precautions [31-37], though additional evaluation is needed to quantify this impact. Finally, with the exception of a desire by some patients for more hands-on examination and in-person interaction, patients felt their quality of care was no different as a result of inpatient telemedicine.

Direct and implied recommendations to improve inpatient telemedicine include increasing audio volume, allowing for a digital "knock" to serve as a warning prior to automatic turn-on, establishing an unobtrusive stable location for the tablet, adding a call-out button to the nursing team, increasing focus on physical touch, improving clinical team introductions, and standardizing patient orientation to the technology (Table 2).

These data are consistent with the past literature, which suggests that providers may overestimate their communication ability [41]. Given that a physician is required to evaluate each inpatient at least once per day, the frequency of physician encounters less than once per day reported by some patient suggests suboptimal communication and role confusion. A strong introduction includes the name, role, and responsibilities of each member of the care team, information that is often not shared even in nondigital clinical settings [42,43]. Training related to the importance of verbal introductions [44,45] and best practices to foster the patient–provider connection in the virtual setting may help mitigate the challenge of role confusion [46]. Clear badges or face cards [38,43,47,48] can also help, but these tools are not available in the digital setting. Instead, facial recognition technology is being explored to solve other problems in health care [49,50] and could be used to automatically display a virtual name badge.

Some patients were also concerned that aspects of their care may be missed owing to the lack of a hands-on examination. The importance of a hands-on physical examination—both for diagnostic and therapeutic purposes—is evidenced in the literature [51,52], but how to best adapt these learnings to patients are under isolation precautions remains an open question. Nurses were reported to use telemedicine more frequently than physicians, while also maintaining a frequent physical presence within the room, which is perhaps unsurprising given that nurses spend approximately 6-fold more time at the patient bedside than physicians in non–COVID-19 settings [53,54]. These COVID-19 data further validate past qualitative work, which suggests that nurses use telemedicine as a bridge to in-person care and "batch" care activities, such that physical assessment, medication delivery, meal delivery, and sanitation protocols all occur in a single room entry [27,55]. The impact of this shift, alongside the reported reduction in in-person physician assessments, on clinical outcomes is an area for future research.

To this end, inpatient telemedicine risks exacerbating pre-existing clinician reliance on technology over hands-on examination. Substitutes for an in-person examination, such as directing patients toward a self-examination with a digital stethoscope [56], pulse oximeter, or other technologies capable of remote transmission of data, even if appropriate from a diagnostic perspective [57], may be less acceptable to patients in the inpatient setting. Additional evaluation is therefore needed to determine the optimal ratio of in-person to virtual encounters.

Table 2. Implied technical and protocol recommendations to improve inpatient telemedicine in accordance with patient interviews.

Patient concern	Possible solutions
Telemedicine technical setup	
Low audio volume	• Improve the audio with more powerful provider-side microphone and patient-side speaker technology
Tablet position is suboptimally angled or too far from the patient	• Stable, unobtrusive "home" for the tablet device within the patient room at an optimal distance and angle; standardized within all patient rooms with telemedicine capabilities
Tablet device automatically turns on without warning	• Announcement of an incoming call with a digital "knock" and a visual and audible countdown prior to automatic web conference turn-on or "entry" into a room
Emergency situation where the patient wants immediate contact with the care team	• Callout button direct to the nurse from the tablet device and the web conferencing system
Care team protocol when using telemedicine	
Desire for physical examination on a regular basis	• Hands-on physical examination by the physician or nursing team with a dedicated stethoscope • Exploration of patient self-exam using enabled devices such as stethoscope with remote transmission capabilities
Poor understanding of who is on the care team and their respective roles when using the web conferencing tool	• Clinical team training emphasizing improved verbal introductions at each virtual encounter • Automatic caption with the name or title based on facial recognition technology
Insufficient patient orientation to tablet use	• Standard orientation to telemedicine, including self-directed exploration

Limitations

Insights from this evaluation are drawn from a small sample of patients within a single institution and therefore cannot be more broadly generalized. These exploratory interviews provide insight into an otherwise difficult-to-reach patient population with currently limited available data, though future work will benefit from expansion to diverse institutions and patient populations. Including patients who required intensive care may be particularly informative; safety and feasibility constraints limited our ability to capture these perspectives in this assessment. In addition, we purposefully sought a diverse set of voices in our sampling protocol, and our final sample overrepresented non-White participants on the basis of local demographics [58]. Increasing sample diversity in terms of languages spoken, given the predominance of English speakers in this sample, may benefit future work. Finally, these qualitative data may also complement and inform ongoing and future quantitative work that explores the impact of inpatient telemedicine on clinician workflows and, by extension, infection control and resource use [59].

Conclusions

Inpatient telemedicine adopted for the purposes of infection control during the COVID-19 pandemic presents a novel use case of the technology, and our understanding of its impact on clinical workflows, patient outcomes, and the patient experience continues to evolve. This qualitative evaluation suggests that while patients still prefer in-person interactions in the hospital setting, inpatient telemedicine is broadly accepted given the need for COVID-19 isolation precautions. Perceived benefits include increased access to the clinical team and reduced anxiety, yet challenges around the technical setup, clinical team introductions, and physical examination remain. Further evaluation is needed to understand the impact on clinical outcomes and the optimal balance between in-person and virtual care in the hospital setting.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Semi-structured patient interview protocol.
[DOCX File , 19 KB - formative_v6i3e32933_app1.docx]

References

1. Hollander JE, Carr BG. Virtually perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]

2. Ohannessian R, Duong TA, Odone A. Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: a call to action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]

3. Vilendrer S, Patel B, Chadwick W, Hwa M, Asch S, Pageler N, et al. Rapid deployment of inpatient telemedicine in response to COVID-19 across three health systems. J Am Med Inform Assoc 2020 Jul 01;27(7):1102-1109 [FREE Full text] [doi: 10.1093/jamia/ocaa077] [Medline: 32495830]

4. Umoren RA, Gray MM, Handley S, Johnson N, Kunimura C, Mietzsch U, et al. In-hospital telehealth supports care for neonatal patients in strict isolation. Am J Perinatol 2020 Jun;37(8):857-860 [FREE Full text] [doi: 10.1055/s-0040-1709687] [Medline: 32268382]

5. Meyer BC, Friedman LS, Payne K, Moore L, Cressler J, Holberg S, et al. Medical undistancing through telemedicine: a model enabling rapid telemedicine deployment in an academic health center during the COVID-19 pandemic. Telemed J E Health 2021 Jun;27(6):625-634. [doi: 10.1089/tmj.2020.0327] [Medline: 33030985]

6. Pilosof NP, Barrett M, Oborn E, Barkai G, Pessach IM, Zimlichman E. Telemedicine implementation in COVID-19 ICU: balancing physical and virtual forms of visibility. HERD 2021 Jul;14(3):34-48 [FREE Full text] [doi: 10.1177/19375867211009225] [Medline: 34075789]

7. Wosik J, Fudim M, Cameron B, Gellad ZF, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Jun 01;27(6):957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]

8. Polinski JM, Barker T, Gagliano N, Sussman A, Brennan TA, Shrank WH. Patients' satisfaction with and preference for telehealth visits. J Gen Intern Med 2016 Mar;31(3):269-275 [FREE Full text] [doi: 10.1007/s11606-015-3489-x] [Medline: 26269131]

9. Aungst LA. Can telemedicine improve triage and patient satisfaction in urgent care settings? J Am Assoc Nurse Pract 2019 Mar;31(3):162-166. [doi: 10.1097/JXX.0000000000000117] [Medline: 30431551]

10. Pruthi S, Stange KJ, Malagrino GD, Chawla KS, LaRusso NF, Kaur JS. Successful implementation of a telemedicine-based counseling program for high-risk patients with breast cancer. Mayo Clin Proc 2013 Jan;88(1):68-73. [doi: 10.1016/j.mayocp.2012.10.015] [Medline: 23274020]

11. Staicu ML, Holly AM, Conn KM, Ramsey A. The use of telemedicine for penicillin allergy skin testing. J Allergy Clin Immunol Pract 2018;6(6):2033-2040. [doi: 10.1016/j.jaip.2018.04.038] [Medline: 29751152]

12. Taylor L, Capling H, Portnoy JM. Administering a telemedicine program. Curr Allergy Asthma Rep 2018 Sep 15;18(11):57. [doi: 10.1007/s11882-018-0812-8] [Medline: 30220060]

13. Kruse CS, Krowski N, Rodriguez B, Tran L, Vela J, Brooks M. Telehealth and patient satisfaction: a systematic review and narrative analysis. BMJ Open 2017 Aug 03;7(8):e016242 [FREE Full text] [doi: 10.1136/bmjopen-2017-016242] [Medline: 28775188]

14. Qiang JK, Marras C. Telemedicine in Parkinson's disease: A patient perspective at a tertiary care centre. Parkinsonism Relat Disord 2015 May;21(5):525-528. [doi: 10.1016/j.parkreldis.2015.02.018] [Medline: 25791380]

15. Hong Z, Li N, Li D, Li J, Li B, Xiong W, et al. Telemedicine during the COVID-19 pandemic: experiences from Western China. J Med Internet Res 2020 May 08;22(5):e19577 [FREE Full text] [doi: 10.2196/19577] [Medline: 32349962]

16. Bokolo AJ. Exploring the adoption of telemedicine and virtual software for care of outpatients during and after COVID-19 pandemic. Ir J Med Sci 2021 Feb;190(1):1-10 [FREE Full text] [doi: 10.1007/s11845-020-02299-z] [Medline: 32642981]

17. Khairat S, Pillai M, Edson B, Gianforcaro R. Evaluating the telehealth experience of patients with COVID-19 symptoms: recommendations on best practices. J Patient Exp 2020 Oct;7(5):665-672 [FREE Full text] [doi: 10.1177/2374373520952975] [Medline: 33294596]

18. Ramaswamy A, Yu M, Drangsholt S, Ng E, Culligan PJ, Schlegel PN, et al. Patient satisfaction with telemedicine during the COVID-19 pandemic: retrospective cohort study. J Med Internet Res 2020 Sep 09;22(9):e20786 [FREE Full text] [doi: 10.2196/20786] [Medline: 32810841]

19. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: Evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]

20. Moeckli J, Gutierrez J, Kaboli PJ. Perceived need and potential applications of a telehospitalist service in rural areas. Telemed J E Health 2021 Jan;27(1):90-95. [doi: 10.1089/tmj.2020.0018] [Medline: 32316876]

21. McSwain SD, Marcin JP. Telemedicine for the care of children in the hospital setting. Pediatr Ann 2014 Feb;43(2):e44-e49. [doi: 10.3928/00904481-20140127-10] [Medline: 24512161]

22. Viers BR, Pruthi S, Rivera ME, O'Neil DA, Gardner MR, Jenkins SM, et al. Are patients willing to engage in telemedicine for their care: a survey of preuse perceptions and acceptance of remote video visits in a urological patient population. Urology 2015 Jun;85(6):1233-1239. [doi: 10.1016/j.urology.2014.12.064] [Medline: 25863832]

23. Portnoy JM, Pandya A, Waller M, Elliott T. Telemedicine and emerging technologies for health care in allergy/immunology. J Allergy Clin Immunol 2020 Feb;145(2):445-454. [doi: 10.1016/j.jaci.2019.12.903] [Medline: 32035604]

24. Keller JJ, Johnson JP, Latour E. Inpatient teledermatology: Diagnostic and therapeutic concordance among a hospitalist, dermatologist, and teledermatologist using store-and-forward teledermatology. J Am Acad Dermatol 2020 May;82(5):1262-1267. [doi: 10.1016/j.jaad.2020.01.030] [Medline: 31972258]

25. Massey PA, McClary K, Zhang AS, Savoie FH, Barton RS. Orthopaedic surgical selection and inpatient paradigms during the coronavirus (COVID-19) pandemic. J Am Acad Orthop Surg 2020 Jun 01;28(11):436-450 [FREE Full text] [doi: 10.5435/JAAOS-D-20-00360] [Medline: 32304401]

26. Griebeler ML, Pantalone KM, Gambino R, Shewmon D, Morrow J, Mendlovic D, et al. The importance of implementing inpatient virtual coverage in an endocrinology practice: lessons learned thus far from the COVID-19 pandemic. Clin Diabetes Endocrinol 2021 Feb 09;7(1):5 [FREE Full text] [doi: 10.1186/s40842-021-00118-7] [Medline: 33557919]

27. Safaeinili N, Vilendrer S, Williamson E, Zhao Z, Brown-Johnson C, Asch SM, et al. Inpatient telemedicine implementation as an infection control response to COVID-19: qualitative process evaluation study. JMIR Form Res 2021 Jun 16;5(6):e26452 [FREE Full text] [doi: 10.2196/26452] [Medline: 34033576]

28. Halabi R, Smith G, Sylwestrzak M, Clay B, Longhurst CA, Lander L. The impact of inpatient telemedicine on personal protective equipment savings during the COVID-19 pandemic: cross-sectional study. J Med Internet Res 2021 May 19;23(5):e28845 [FREE Full text] [doi: 10.2196/28845] [Medline: 33945494]

29. Bulik RJ. Human factors in primary care telemedicine encounters. J Telemed Telecare 2008;14(4):169-172. [doi: 10.1258/jtt.2007.007041] [Medline: 18534948]

30. Gruppen LD, Wolf FM, Billi JE. Information gathering and integration as sources of error in diagnostic decision making. Med Decis Making 1991;11(4):233-239. [doi: 10.1177/0272989X9101100401] [Medline: 1766327]

31. Barratt RL, Shaban R, Moyle W. Patient experience of source isolation: lessons for clinical practice. Contemp Nurse 2011 Oct;39(2):180-193. [doi: 10.5172/conu.2011.180] [Medline: 22551431]

32. Abad C, Fearday A, Safdar N. Adverse effects of isolation in hospitalised patients: a systematic review. J Hosp Infect 2010 Oct;76(2):97-102 [FREE Full text] [doi: 10.1016/j.jhin.2010.04.027] [Medline: 20619929]

33. Siddiqui ZK, Conway SJ, Abusamaan M, Bertram A, Berry SA, Allen L, et al. Patient isolation for infection control and patient experience. Infect Control Hosp Epidemiol 2019 Feb;40(2):194-199. [doi: 10.1017/ice.2018.324] [Medline: 30560748]

34. Nair R, Perencevich E, Goto M, Livorsi D, Balkenende E, Kiscaden E, et al. Patient care experience with utilization of isolation precautions: systematic literature review and meta-analysis. Clin Microbiol Infect 2020 Jun;26(6):684-695 [FREE Full text] [doi: 10.1016/j.cmi.2020.01.022] [Medline: 32006691]

35. Sun N, Wei L, Wang H, Wang X, Gao M, Hu X, et al. Qualitative study of the psychological experience of COVID-19 patients during hospitalization. J Affect Disord 2021 Jan 01;278:15-22 [FREE Full text] [doi: 10.1016/j.jad.2020.08.040] [Medline: 32949869]

36. Vottero B, Rittenmeyer L. The hospitalised patients' experience of being in protective/source isolation: A systematic review of qualitative evidence. JBI Libr Syst Rev 2012;10(16):935-976. [doi: 10.11124/01938924-201210160-00001] [Medline: 27820463]

37. Fan PEM, Aloweni F, Lim SH, Ang SY, Perera K, Quek AH, et al. Needs and concerns of patients in isolation care units - learnings from COVID-19: A reflection. World J Clin Cases 2020 May 26;8(10):1763-1766 [FREE Full text] [doi: 10.12998/wjcc.v8.i10.1763] [Medline: 32518768]

38. Brown-Johnson C, Vilendrer S, Heffernan MB, Winter S, Khong T, Reidy J, et al. PPE portraits-a way to humanize personal protective equipment. J Gen Intern Med 2020 Jul;35(7):2240-2242 [FREE Full text] [doi: 10.1007/s11606-020-05875-2] [Medline: 32410125]

39. Creswell J, Poth C. Qualitative Inquiry and Research Design Choosing Among Five Approaches (4th edition). Thousand Oaks, CA: Sage Publications; 2016.

40. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]

41. Sethuraman KR. Doctor-Patient Communication: An Overview. In: Communication Skills in Clinical Practice (Doctor-Patient Communication). New Delhi: Jaypee Digital; 2001:1-11.

42. Dalal AK, Schnipper JL. Care team identification in the electronic health record: A critical first step for patient-centered communication. J Hosp Med 2016 May;11(5):381-385. [doi: 10.1002/jhm.2542] [Medline: 26762584]

43. Hickerton BC, Fitzgerald DJ, Perry E, De Bolla AR. The interpretability of doctor identification badges in UK hospitals: a survey of nurses and patients. BMJ Qual Saf 2014 Jul;23(7):543-547. [doi: 10.1136/bmjqs-2013-002445] [Medline: 24599731]

44. Allenbaugh J, Spagnoletti CL, Rack L, Rubio D, Corbelli J. Health literacy and clear bedside communication: a curricular intervention for internal medicine physicians and medicine nurses. MedEdPORTAL 2019 Jan 18;15:10795 [FREE Full text] [doi: 10.15766/mep_2374-8265.10795] [Medline: 30800995]

45. Kelly MM, Xie A, Carayon P, DuBenske LL, Ehlenbach ML, Cox ED. Strategies for improving family engagement during family-centered rounds. J Hosp Med 2013 Apr;8(4):201-207 [FREE Full text] [doi: 10.1002/jhm.2022] [Medline: 23468375]

46. Zulman DM, Verghese A. Virtual care, telemedicine visits, and real connection in the era of COVID-19: unforeseen opportunity in the face of adversity. JAMA 2021 Feb 02;325(5):437-438. [doi: 10.1001/jama.2020.27304] [Medline: 33528520]

47. Lill MM, Wilkinson TJ. Judging a book by its cover: descriptive survey of patients' preferences for doctors' appearance and mode of address. BMJ 2005 Dec 24;331(7531):1524-1527 [FREE Full text] [doi: 10.1136/bmj.331.7531.1524] [Medline: 16373739]

48. Arora VM, Schaninger C, D'Arcy M, Johnson JK, Humphrey HJ, Woodruff JN, et al. Improving inpatients' identification of their doctors: use of FACE cards. Jt Comm J Qual Patient Saf 2009 Dec;35(12):613-619 [FREE Full text] [doi: 10.1016/s1553-7250(09)35086-2] [Medline: 20043501]

49. Chanda A, Chatterjee S. Predicting Obesity Using Facial Pictures during COVID-19 Pandemic. Biomed Res Int 2021;2021:6696357 [FREE Full text] [doi: 10.1155/2021/6696357] [Medline: 33778081]

50. Postma-Nilsenová M, Postma E, Tates K. Automatic detection of confusion in elderly users of a web-based health instruction video. Telemed J E Health 2015 Jun;21(6):514-519. [doi: 10.1089/tmj.2014.0061] [Medline: 25844904]

51. While A. Touch: knowledge and considerations for nursing practice. Br J Community Nurs 2021 Apr 02;26(4):190-194. [doi: 10.12968/bjcn.2021.26.4.190] [Medline: 33797963]

52. Bruhn JG. The doctor's touch: tactile communication in the doctor-patient relationship. South Med J 1978 Dec;71(12):1469-1473. [doi: 10.1097/00007611-197812000-00008] [Medline: 364668]

53. Li RC, Marafino BJ, Nielsen D, Baiocchi M, Shieh L. Assessment of a real-time locator system to identify physician and nurse work locations. JAMA Netw Open 2020 Feb 05;3(2):e1920352 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.20352] [Medline: 32022876]

54. Mitchell M, Lavenberg J, Trotta R, Umscheid C. Hourly rounding to improve nursing responsiveness: a systematic review. J Nurs Adm 2014 Sep;44(9):462-472 [FREE Full text] [doi: 10.1097/NNA.0000000000000101] [Medline: 25148400]

55. Schroeder K, Norful AA, Travers J, Aliyu S. Nursing perspectives on care delivery during the early stages of the covid-19 pandemic: A qualitative study. Int J Nurs Stud Adv 2020 Nov;2:100006 [FREE Full text] [doi: 10.1016/j.ijnsa.2020.100006] [Medline: 32864632]

56. Lakhe A, Sodhi I, Warrier J, Sinha V. Development of digital stethoscope for telemedicine. J Med Eng Technol 2016;40(1):20-24. [doi: 10.3109/03091902.2015.1116633] [Medline: 26728637]

57. Portnoy JM, Waller M, De Lurgio S, Dinakar C. Telemedicine is as effective as in-person visits for patients with asthma. Ann Allergy Asthma Immunol 2016 Sep;117(3):241-245. [doi: 10.1016/j.anai.2016.07.012] [Medline: 27613456]

58. San Francisco Bay Area. Bay Area Census. URL: http://www.bayareacensus.ca.gov/bayarea.htm [accessed 2020-12-16]

59. Patel B, Vilendrer S, Kling SMR, Brown I, Ribeira R, Eisenberg M, et al. Using a real-time locating system to evaluate the impact of telemedicine in an emergency department during COVID-19: observational study. J Med Internet Res 2021 Jul 26;23(7):e29240 [FREE Full text] [doi: 10.2196/29240] [Medline: 34236993]

Abbreviations

PPE: personal protective equipment

XSL•FO

RenderX

Original Paper

Understanding Engagement and the Potential Impact of an Electronic Drug Repository: Multi-Methods Study

Charlene Soobiah[1,2], HBSc; Michelle Phung[1], MSc; Mina Tadrous[1,3], MSc, PharmD, PhD; Trevor Jamieson[1,4], MD, MBI; R Sacha Bhatia[1], MD, MBA; Laura Desveaux[1,2], MScPT, PhD

[1]Institute for Health System Solutions and Virtual Care, Women's College Hospital, Toronto, ON, Canada

[2]Institute for Health Policy, Management & Evaluation, University of Toronto, Toronto, ON, Canada

[3]Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, ON, Canada

[4]Unity Health Toronto, Toronto, ON, Canada

Corresponding Author:
Laura Desveaux, MScPT, PhD
Institute for Health System Solutions and Virtual Care
Women's College Hospital
76 Grenville St
Toronto, ON, M5S 1B2
Canada
Phone: 1 (416) 323 6400
Email: laura.desveaux@thp.ca

Abstract

Background: Centralized drug repositories can reduce adverse events and inappropriate prescriptions by enabling access to dispensed medication data at the point of care; however, how they achieve this goal is largely unknown.

Objective: This study aims to understand the perceived clinical value; the barriers to and enablers of adoption; and the clinician groups for which a provincial, centralized drug repository may provide the most benefit.

Methods: A mixed methods approach, including a web-based survey and semistructured interviews, was used. Participants were clinicians (eg, nurses, physicians, and pharmacists) in Ontario who were eligible to use the digital health drug repository (DHDR), irrespective of actual use. Survey data were ranked on a 7-point adjectival scale and analyzed using descriptive statistics, and interviews were analyzed using qualitative descriptions.

Results: Of the 161 survey respondents, only 40 (24.8%) actively used the DHDR. Perceptions of the utility of the DHDR were neutral (mean scores ranged from 4.11 to 4.76). Of the 75.2% (121/161) who did not use the DHDR, 97.5% (118/121) rated access to medication information (eg, dose, strength, and frequency) as important. Reasons for not using the DHDR included the cumbersome access process and the perception that available data were incomplete or inaccurate. Of the 33 interviews completed, 26 (79%) were active DHDR users. The DHDR was a satisfactory source of secondary information; however, the absence of medication instructions and prescribed medications (which were not dispensed) limited its ability to provide a comprehensive profile to meaningfully support clinical decision-making.

Conclusions: Digital drug repositories must be adjusted to align with the clinician's needs to provide value. Ensuring integration with point-of-care systems, comprehensive clinical data, and streamlined onboarding processes would optimize clinically meaningful use. The electronic provision of accessible drug information to providers across health care settings has the potential to improve efficiency and reduce medication errors.

XSL•FO
RenderX

Introduction

Background

Pharmacological management of chronic diseases and acute illness is common, with over 41% of Canadians reporting using at least one prescription drug [1]. The lack of comprehensive medication information at the point of care can increase the likelihood of an adverse drug event (ADE) [2,3]. A meta-analysis of 46,000 patient records estimated that 20% of hospital inpatients in high income countries have at least one ADE, of which 32% are deemed preventable [4]. In community settings, 5% of patients experience at least one ADE in their lifetime [5], which translates into 16 million Americans or 2 million Canadians based on current population estimates [6,7]. ADEs are often caused by preventable prescribing errors (ie, wrong dose or therapeutic choice) and ineffective monitoring of pharmaceutical care [5,8], costing health care systems an estimated CAD $1.1 billion per year (US $0.89 billion) [9].

Medication reconciliation reduces the likelihood of ADEs [10] and involves a collaborative effort between clinicians and patients to establish a best possible medication history (BPMH) to improve medication safety [11,12]. Medication histories reported by patients during the BPMH process are often verified using secondary sources of information, such as pharmacy or discharge records [13]; however, this process is time consuming and cumbersome at the point of care.

There are several examples of national and international drug repositories, such as the British Columbia Pharmanet [14] or Sweden's National Prescription Drug record [15], which typically contain dispensed medication information, such as the drug name, the dose, allergy, and intolerance information, which can be valuable when constructing a BPMH. A few drug repositories also include information on prescribed medications and private insurance claims to facilitate decision-making at the point of care [16,17]. Access to a centralized drug repository can support the BPMH process and, in turn, reduce inappropriate prescribing [18], improve medication adherence [19], and reduce health care costs [20,21].

Objectives

To achieve these aims, the Ontario Ministry of Health implemented the digital health drug repository (DHDR). The DHDR contains information on publicly funded drugs in Ontario, pharmacy services, and dispensed monitored drugs (ie, narcotics and controlled substances). The DHDR was developed by a provincial agency responsible for creating a public electronic health record system (eHealth Ontario), and access is enabled through one of two provincial clinical viewers (*ClinicalConnect* and *Connecting Ontario*), which provides access to a range of health system data assets. The DHDR was implemented in 2016 and had over 150,000 registered users across over 546 sites at the start of this study (November 2018). Eligible users of provincial viewers include clinicians who require patient medical information as part of the provision of care (eg, physicians, nurses, and pharmacists). Use statistics estimate that approximately 3000 users access the DHDR daily; however, the drivers of engagement and whether the DHDR is achieving its objectives remain unclear.

The overarching aim of this work is to examine the use, drivers of engagement, perceived value, and potential impact of the DHDR. The specific objectives are to (1) understand the perceived clinical value of accessing medication history via the DHDR, (2) identify enablers of and barriers to adoption among eligible users, and (3) understand for which clinician groups the DHDR has current or future potential clinical value and how that value is (or might be) realized.

Methods

Overview

A multi-method approach was used to elicit feedback from users and nonusers of the DHDR and included a cross-sectional electronic survey and semistructured interviews. The approach allowed us to simultaneously collect a breadth of responses from users and nonusers in the cross-sectional survey, whereas the interviews explored more fulsomely individual experiences with the DHDR. Interviews and the survey were conducted concurrently to maximize research efforts, and the results were triangulated and interpreted simultaneously (ie, the qualitative findings were used to help understand the survey findings and vice versa). Informed consent was obtained electronically for participants completing the survey, and verbal consent was obtained for those participating in interviews.

Ethics Approval

Research ethics approval was obtained from the Women's College Hospital using the assessment process for quality improvement projects (WCH APQIP REB #2019-0038).

Study Setting

In Ontario, some prescription drugs were publicly covered by the Drugs and Devices Division of the Ministry of Health (formerly Ontario Public Drug Programs). These include the Ontario Drug Benefit Program for residents aged >65 years, the Trillium Drug Program for residents with high medication costs in relation to household income, and other specialized drug programs for those with complex conditions, such as cancer, inherited metabolic disorders, and those receiving home care [22]. In 2013, over 2.8 million, or approximately 20% of residents, received benefits through these programs to reduce the cost of medications [22]. In addition, the DHDR incorporates the provincial Narcotics Monitoring System, which captures dispensed narcotics (ie, opioids) and controlled substances (ie, methylphenidates, benzodiazepines, and barbiturates) for all residents [23]. The DHDR includes dispensed medication information from the Drugs and Devices Division Programs and the Narcotics Monitoring System; however, it does not include medications paid through private insurance companies; over-the-counter medications; or general purchase medications (eg, acetaminophen), supplements, or medication samples (ie, novel anticoagulants) provided by pharmaceutical companies.

To centralize access to provincial digital assets, the DHDR was embedded into 2 pre-existing clinical viewing portals. These portals allow providers access to a range of patient-level information, including diagnostic imaging reports, dispensed medications, laboratory results, hospital visits, and home and community care information (ie, referral details, risk

assessments, and care plans) in Ontario. Although participants' perceptions of the DHDR were influenced by their experience with the clinical viewer, exploring its functionality was beyond the scope of this study.

Participant Recruitment

Clinicians from all sectors of the health care system were eligible for participation in the survey or interviews, provided they were clinicians who required access to electronic patient records as part of the provision of clinical care. This group includes physicians, specialists, nurses, pharmacists, and other allied health providers employed at health organizations. Allied health professionals employed in the private sector such as physiotherapists, occupational therapists, and paramedics were not eligible.

Recruitment for the survey and interviews was conducted concurrently using a multipronged approach, as no mechanism existed to identify or contact current users directly. Our first strategy involved recruiting participants through Local Registration Authorities (LRAs) working with eHealth Ontario. LRAs are individuals nominated by their organization or site to train users on clinical viewers (ie, *Connecting Ontario* and *ClinicalConnect*) and its connected repositories such as the DHDR. Using aggregate use data obtained from eHealth Ontario, we stratified the data based on the type of health care setting (ie, acute care, long-term care, and community care) and region (eg, Southwest, Northeast, and Central Ontario). Our goal was to recruit from sites with a larger pool of users to maximize the potential for recruitment. Thus, sites with fewer than 20 registered users were excluded (n=308). LRAs were offered a modest honorarium to acknowledge their efforts, and we were limited to include a maximum of 24 sites based on our funding. These sites were selected using a random number generator of the included sites with >20 registered users (n=112). The LRAs from the 24 targeted sites were then asked to distribute recruitment emails.

As the survey was constructed for the purpose of this study, there was no sample size calculation behind our target recruitment. Overall, the selected target sites had approximately 6794 active users and over 15,000 authorized users. We anticipated the LRAs would send the link to the survey to at least 50% (3397/6794) of active users, of which we estimate 30% (1019/3397) would open the email and 20% (203/1019) of which will complete the survey [24,25].

To increase participation, our secondary strategy involved recruitment through the DHDR Clinical Working Group, a group of clinicians who are actively engaged in digital health solutions at their respective organizations and advise eHealth Ontario on how to improve the DHDR. The DHDR working group members circulated a recruitment email detailing the project, contact information of our study team, and a link to the electronic survey through their internal networks. Finally, we recruited individuals through our internal networks and social media outlets (Twitter and LinkedIn). Participants who were interested would click the link to the survey or contact the study coordinator to participate in an interview.

Data Collection

Cross-sectional Survey

A web-based cross-sectional survey targeted authorized users and nonauthorized potential users of the DHDR to understand their knowledge related to the DHDR and how they use its data where applicable. Nonauthorized users are individuals who are eligible for DHDR access but have not been issued secure credentials for access to the repository (ie, still not registered for access through eHealth Ontario via a clinical viewer).

Survey items were informed by previous Ministry of Health benefits and evaluation reports and past surveys conducted by eHealth Ontario. The questions were further modified and reviewed by stakeholders at the Ministry of Health and eHealth Ontario to ensure sufficient alignment to inform decision-making. In addition, to ensure we included questions relevant to our evaluation, face validity in relation to our study objective was assessed using the Clinical Sensibility Questionnaire (Multimedia Appendix 1) [26]. Three stakeholders from eHealth Ontario, 4 clinicians, and 1 researcher (outside of the research team) from Women's College Hospital reviewed the items for inclusion. A total of 10 questions were removed to eliminate redundancy, and several questions were rephrased for clarity based on feedback.

Survey items included a demographic questionnaire followed by questions on the perceptions of the DHDR (Multimedia Appendix 2). DHDR users rated their experience across four key domains: (1) usefulness, (2) quality of data, (3) training, and (4) satisfaction. Participants also rated the perceived value of the 14 data elements currently contained in the repository (eg, strength, dose, and therapeutic class) and their perceptions of the overall value and impact of the DHDR. Finally, the participants ranked the importance of enhancing the DHDR with specific questions currently under consideration for integration: (1) the comprehensive inclusion of all prescribed medications (ie, the addition of those that are not dispensed), (2) privately paid drugs (ie, private insurance or out-of-pocket claims), and (3) additional clinical data (eg, tolerances or allergies). Nonusers of the DHDR were asked to rate their perceptions of the value of centralized repository access and its potential impact. Further questions using open textboxes were used to elicit the reasons why participants did not use the DHDR and the resources used to develop a BPMH.

Survey items were rated on a 7-point adjectival scale with anchors ranging from *1* (*strongly disagree*) to *7* (*strongly agree*) or *1* (*not at all important/valuable*) to *7* (*extremely important/valuable*). The survey was administered and managed on the web using the Research Electronic Data Capture (REDCap) tool hosted at WCH, in Toronto, Ontario [27,28]. It is a secure web-based software platform designed to support data capture for research studies. Participants who completed the survey had the opportunity to enter into a draw for 1 of the 3 CAD $100 (US $79.46) gift cards.

Semistructured Interviews

Interviews were conducted with users and nonusers of the DHDR by trained qualitative research staff with no prior relationship with the study participants (CS and MP), under the

supervision of an experienced qualitative researcher (LD). The interviews focused on mechanisms for accessing the DHDR (ie, type of clinical viewer used), features and functions of the repository, perceptions of the data elements contained in the repository, barriers to or enablers of adoption, and potential impact on clinical workflow and health outcomes (Multimedia Appendix 3). Questions were tailored by user type (users and nonusers of the DHDR) and included a demographic questionnaire. Before conducting interviews, the interview guides were reviewed by an experienced qualitative researcher (LD), a pharmacoepidemiologist (MT), and a researcher (external to the research team with expertise in digital health evaluations). The interview questions were derived directly from the research aims, future features, and functionality of the DHDR. The interviews were conducted in person or over the phone, according to the participant's preference. The interviews were audio-recorded, transcribed, and anonymized by an independent third party.

Analysis

Cross-sectional Survey

Participants were able to skip questions or end the survey early as the survey did not force a response. These incomplete responses were included in the analysis, resulting in different denominators across different questions, and participant responses were aggregated and descriptively analyzed. Items rated on the 7-point adjectival scale using the anchors *strongly disagree* to *strongly agree* were interpreted as disagreement, neutral, and agreement by consolidating responses. Negative statements in the 4 key domains were reverse-scored, and the overall means were calculated for each domain. All the other survey items were aggregated. Subgroup analyses were conducted to assess the impact of gender and user and nonuser perceptions of the impact and value of the DHDR. Differences in gender and user and nonuser responses were evaluated using the Wilcoxon Rank-Sum test. Survey data are presented as means and SD, and statistical significance was considered at the 0.05 level. All analyses were conducted in the R statistical program using the *ggpubr* and *MASS* packages [29].

Semistructured Interviews

Interviews were analyzed using qualitative description, a paradigm that seeks to create an understanding of a phenomenon by accessing the meanings that participants ascribe to it [30,31]. Two coders (CS and MP) independently and inductively coded the first three interviews, after which they met to discuss the data, achieve coding alignment, and establish a codebook. One coder (MP) deductively applied the codebook to the remaining interviews, creating inductive codes as new data emerged. New codes were discussed iteratively (CS and MP), and the second coder (CS) reviewed a random subsample of 3 additional interviews to evaluate consistency. Interviews continued until thematic saturation was achieved (ie, no new themes were identified), at which point 3 additional interviews were conducted to confirm saturation [32]. Two team members (CS and MP) then generated preliminary themes that were discussed with a senior scientist (LD) and a pharmacoepidemiologist (MT). Themes were further refined based on group discussions and finalized once a consensus was achieved. Qualitative data were analyzed using NVivo (version 12 Plus; QSR International) [33].

Results

Overview

The survey and interview data were collected between May and August 2019. The results are first presented for surveys, followed by the data obtained from the interviews.

Survey Results

Of the 161 participants who completed the survey, 80.7% (130) were predominantly female and represented a range of health professions (see Table 1 for demographic characteristics). Approximately 32% (53/161) were not using a provincial viewer for patient information, only 24% (38/161) indicated that they were using the DHDR, and 44% (70/161) indicated they were not using the DHDR. Survey validation findings are detailed in Multimedia Appendix 4 [34-36]. Across the key experience domains, the coefficients were negatively correlated, suggesting that the survey items within each section were slightly convergent; however, the findings were not statistically significant.

Table 1. Survey participant demographics (N=161).

Demographic attribute	Values, n (%)
Sex	
Male	30 (18.6)
Female	131 (81.3)
Age (years)	
18-34	52 (32.2)
35-49	42 (26)
50-64	65 (40.3)
≥65	1 (0.6)
Not reported	1 (0.6)
Health care setting	
Acute care	93 (57.7)
Primary care	24 (14.9)
Community care	34 (21.1)
Long-term care	2 (1.2)
Other	9 (5.4)
Primary occupation	
Physician	25 (15.5)
Nurse	52 (32.2)
Pharmacist	45 (27.9)
Allied health professional	9 (5.5)
Support personnel	7 (4.2)
Other[a]	23 (14.2)
Primary source of clinical information	
Hospital information system	68 (42.2)
Electronic medical record	38 (23.46)
Client health and related information system	10 (6.2)
Paper records	22 (13.6)
Other[b]	21 (13)
Not reported	2 (1.8)
Provincial viewer	
ClinicalConnect	82 (50.9)
Connecting Ontario	25 (15.5)
None of the above	53 (31.9)
Not reported	1 (0.6)
Used DHDR[c]	
No	123 (76.3)
Yes	38 (23.6)

[a]Primary occupations include pharmacies and medical technicians.

[b]The primary source of clinical information includes Meditech, Medstracker, Kroll information systems, and community pharmacy systems.

[c]DHDR: Digital Health Drug Repository.

DHDR Users

Of the 40 users who reported using the DHDR, 29 were female (73%), 25 were working in acute care (62%), and 27 were located in an urban setting (68%; Multimedia Appendix 5). Most had accessed the DHDR within the last 6 months, although the frequency of access varied (Multimedia Appendix 6).

Across all experience domains (ie, usefulness, quality of data, training received, and overall satisfaction), the average participant response trended toward neutral (neither agree nor disagree), with notable variability in responses (Figure 1). Differences were negligible between male and female respondents; however, female respondents (mean score 4.83, SD 1.61) perceived the DHDR to fit with the workflow to a greater degree than their male counterparts (mean score 2.73, SD 1.85; P=.003; Multimedia Appendix 7).

Figure 1. Mean scores and SDs for digital health drug repository (DHDR) experience among survey respondents.
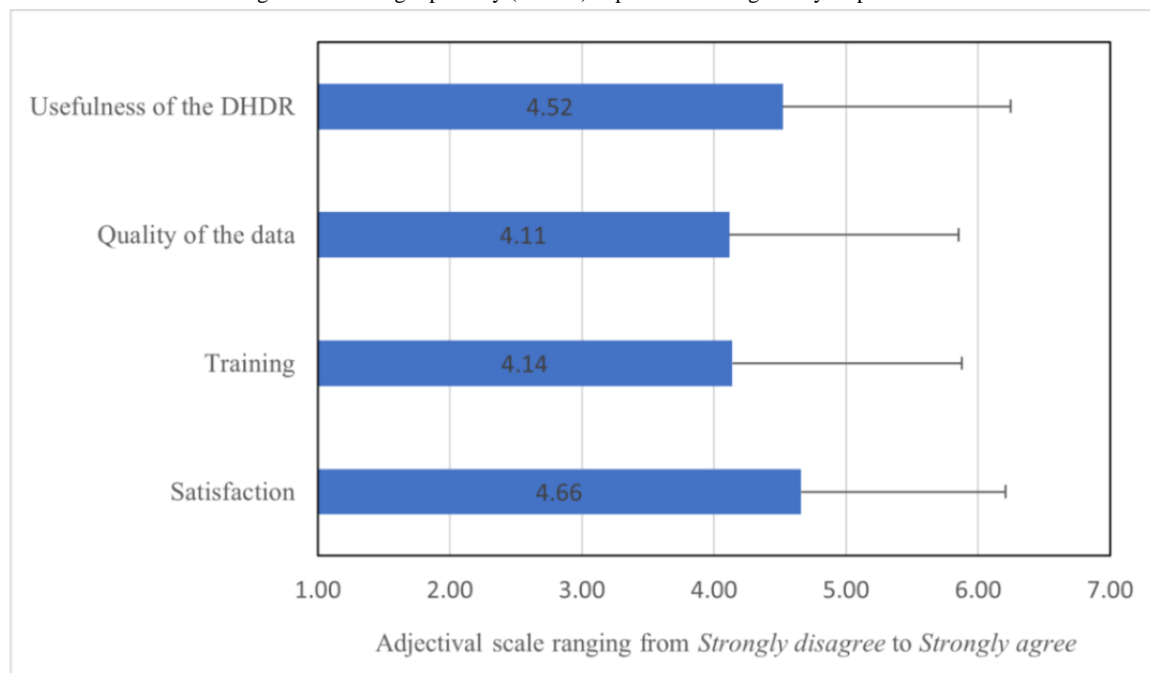
Of the 14 data elements currently included in the DHDR (Multimedia Appendix 8), DHDR users rated 9 elements as necessary for the development of a BPMH (with a mean rating ≥5, and SD >1.90). In terms of perceived importance, these included the strength of dose (mean 6.41, SD 0.82), generic name of the drug (mean 6.07, SD 1.24), quantity of medication dispensed (mean 5.97, SD 1.13), prescriber contact information (mean 5.90, SD 1.20), estimated supply (mean 5.87, SD 1.11), dosage form (mean 5.87, SD 1.32), dispense date (mean 5.85, SD 1.14), pharmacy contact information (mean 5.77, SD 1.47), and prescription count (mean 5.45, SD 1.36; Multimedia Appendix 8). Participants expressed a desire for private insurance claims for dispensed medications, medication instructions (eg, 50 mg twice daily), and explicit discontinuations in the DHDR to facilitate a comprehensive profile of medication history.

On average, participants perceived having access to all prescribed medications, dispensed medications, private insurance claims, additional clinical data, and integration of the DHDR into current point-of-care systems as moderately valuable, with average scores ranging from 5.90 to 6.32 (Multimedia Appendix 9). However, participants were neutral in their perceptions of the perceived impact of the DHDR on reducing ADEs, improving patient outcomes, and reducing communication with other providers, with average scores ranging from 4.76 to 4.21 (Multimedia Appendix 10).

DHDR Nonusers

A total of 123 participants stated that they were not using the DHDR and had a demographic profile similar to that of the DHDR user group (Multimedia Appendix 5). Most respondents (99/123, 80.5%) had not heard of the DHDR, and 89.4% (110/123) were not familiar with what the DHDR did. Surprisingly, 63.4% (78/123) had access to a provincial viewer, which contained the DHDR, but did not use the DHDR itself (Multimedia Appendix 5). Comments reported in the open textboxes suggest that some DHDR nonusers perceived that there was limited clinically meaningful information in the repository, whereas others believed that the records were incomplete, preferring instead to rely on pharmacy records. Those who did not have access to a provincial viewer cited a lack of availability or the tedious registration process as barriers to use.

Comparison of DHDR Users and Nonusers

Overall, users and nonusers of the DHDR expressed the value in the ability to access all prescribed medications, dispensed medications, privately funded medications, and additional clinical data at the point of care when developing a BPMH (Figure 2), with no significant differences between the groups. All participants agreed that a centralized repository would improve patient outcomes, reduce unnecessary communication between clinicians, and reduce ADEs (Figure 3).

Figure 2. Perceptions of the value of accessing information at the point of care among survey respondents. Bars represent mean scores and error bars represent SDs. DHDR: Digital Health Drug Repository.
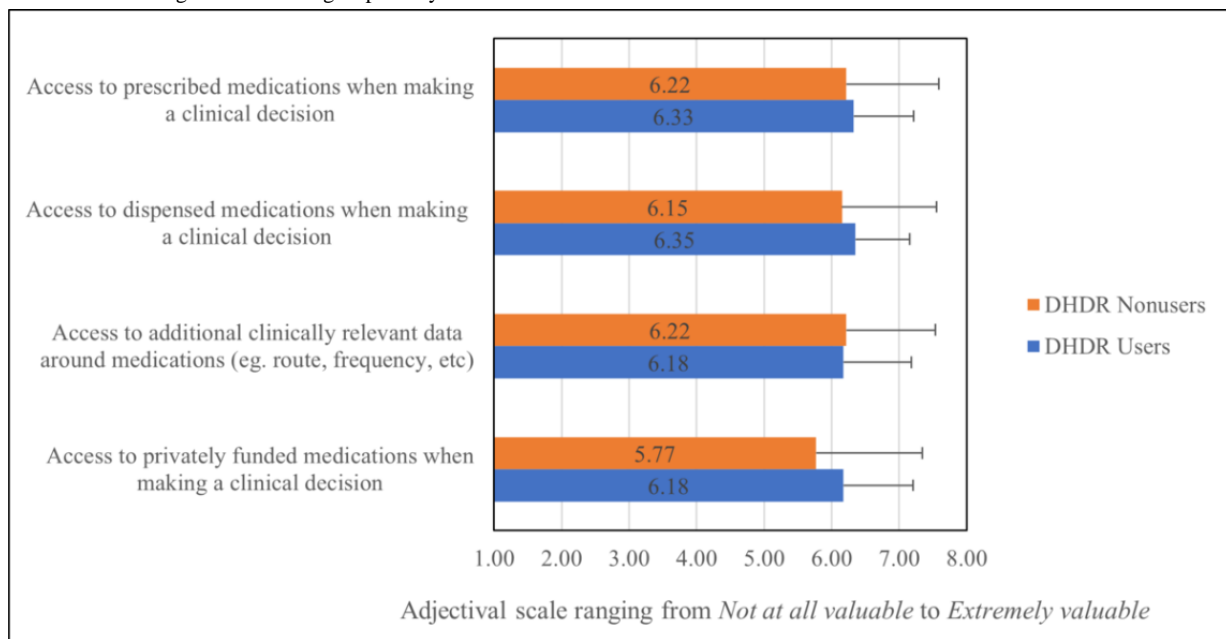
Figure 3. Perceptions of the impact of the digital health drug repository (DHDR) among survey respondents. Bars represent mean scores and error bars represent SDs.
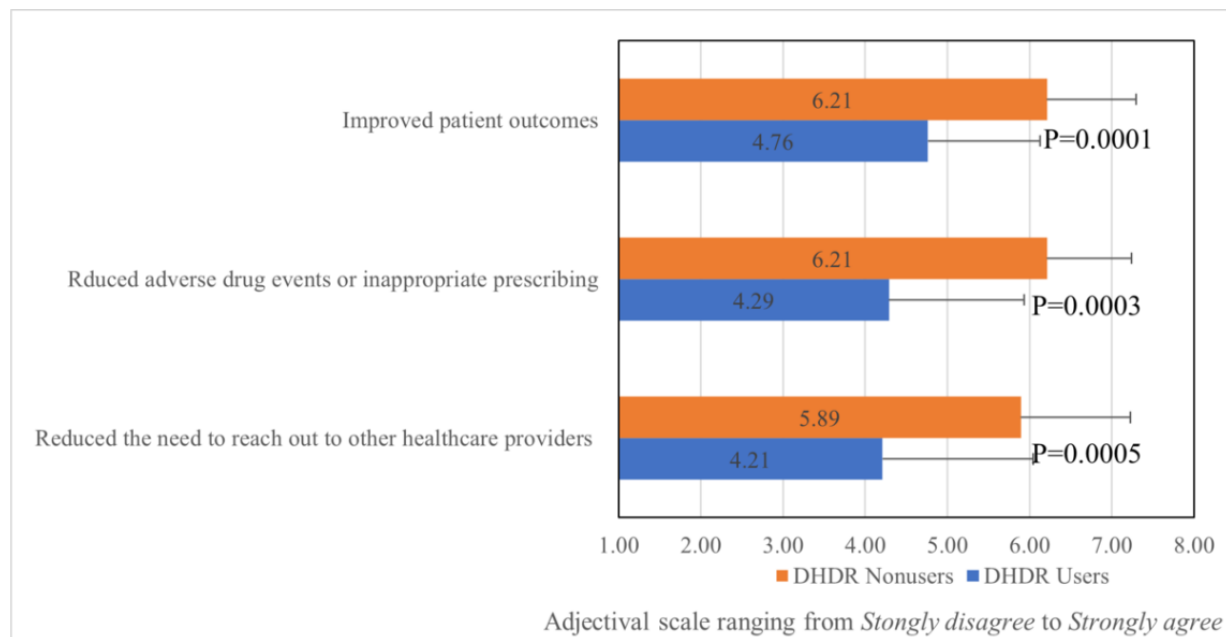
Interviews

A total of 33 clinicians were interviewed between May 13 and August 1, 2019. The average length of the interviews was 25 minutes (range 7-53 minutes). Interviewees had a similar demographic profile to that of the survey respondents, apart from a greater gender balance (Table 2). Of the 33 participants, 26 (79%) had access to a provincial viewer and had accessed the DHDR at least once, and 7 (21%) participants neither had access nor wanted to pursue the process to get access because

they believed that they were ineligible for access or were concerned about the administrative process to obtain access. Moreover, nonusers were generally satisfied with current resources, such as calling pharmacists or health care provider to obtain medication information. Themes centered around user experience, training and onboarding, resources used for BPMH, clinical use cases, impact on clinical workflow, lack of awareness of the DHDR, and perceived impact. Additional quotes supporting key themes can be found in Multimedia Appendix 11.

Table 2. Demographics of interviewed participants (N=33).

Demographic attribute	Participants, n (%)
Gender	
Women	20 (61)
Men	13 (39)
Location	
Urban	23 (70)
Rural	6 (18)
Both	4 (12)
Health care setting	
Acute care	15 (46)
Community care	10 (33)
Primary care	5 (15)
Long-term care	1 (3)
Other[a]	2 (7)
Primary occupation	
Pharmacist	15 (46)
Physician	13 (39)
Other[b]	5 (15)
Provincial clinical viewer	
Clinical connect	13 (39)
Connecting Ontario	13 (39)
No access	7 (21)

[a]Multiple settings (ie, acute and primary care) in the health care setting.

[b]Includes occupational therapists, pharmacy technicians, medical laboratories, and ultrasound technologists in a primary occupation.

User Experience

Overall, 26 participants accessed the repository a few times a month and were generally satisfied with the available data. Participants found the DHDR easy to use and adaptable to their clinical needs (eg, the ability to filter medication histories over a customizable time). Clinicians valued the basic medication information contained in the repository such as drug name, dosage form, and contact information of the prescriber and dispensing pharmacy as secondary mechanisms to validate BPMH.

Participants felt that the DHDR did not fully achieve its intended role as it did not capture all medication information for Ontario residents (ie, only those covered), restricting its ability to provide a complete and comprehensive understanding of a patient's medication history. Relatedly, notable information was missing from the DHDR, including medication instructions, alternative medication names, drug discontinuations, a comprehensive list of prescribed medications (ie, including those that were not dispensed), and private insurance claims for dispensed medications. None of the participants used the DHDR as the primary source of medication information and only used when it was not possible to call the pharmacy or family physician (ie, after hours). Methods that clinicians used to conduct a BPMH

included reviewing the patient's health record, conducting a patient interview, confirming medications with a pharmacist or family physician, and using other resources such as electronic medical records.

Participants were able to identify issues with medications, such as potential drug interactions, ADEs, or duplicate prescriptions because of the DHDR. Some participants expressed that not having access to this information generates unnecessary costs to the health care system because of duplicate prescriptions and risk of complications.

Onboarding and Training

Obtaining access to provincial viewers (within which the DHDR is contained) presented challenges for participants working in community or long-term care settings. Those working in acute care settings gained access through their organization and delays were relatively small (ie, a few days to weeks). For those outside of acute care, the process of obtaining access was tedious and lengthy, with some participants reporting waiting periods to gain access ranging from 2.5 to 18 months. Delays were often the result of security and privacy assessments or waiting for a response to inquiries during registration. The process of obtaining access to the DHDR for a large hospital was the same as that for a small community practice or a local pharmacy, and

participants were often frustrated at the process. Out-of-pocket overhead costs and resources required to achieve provincial viewer access were a concern to community providers, which served as a deterrent to some:

> You basically go on the website and then you look at the link that talks about getting access for your clinic. Then you contact somebody and they usually will take their time to get back to you, but within maybe a month they'll get back to you and they'll put you on the waiting list and you'll wait another month and then they'll call you back to arrange a time. Kind of works like that. Then, they'll come to your clinic to give you a little bit of orientation and then they'll set it up. [P5]

Clinical Use Cases

Participants agreed that having access to comprehensive medication information at the point of care is valuable for clinical decision-making. Access to medication information was highlighted as particularly useful when conducting a BPMH for geriatric patients, complex patients, and those at high risk for readmission (eg, patients with chronic obstructive pulmonary disorder). The DHDR was particularly valuable in emergency departments at night and during surgical consultations when the ability to contact pharmacies and family physicians was either not possible or limited. Among family physicians, the DHDR was primarily used to support clinical decisions related to antibiotic or narcotic prescriptions for new patients.

Factors Limiting Impact

At times, clinicians with access to the DHDR found integration into their clinical workflow challenging. Inefficiencies were created by the need to log out of their usual point-of-care system to log on to the provincial viewer to access the repository. Both DHDR users and nonusers expressed a desire to have centralized repositories integrated within their point-of-care systems to minimize disruptions to the workflow:

> As you can imagine, somebody...is doing a consult, they're opening up their computer, they're looking through [their EMR]...But then, in order to get to ConnectingOntario you have to actually open up a different window...you actually have to come out of [the EMR] to load up another window, which takes you away from what you were doing before...That takes time. It takes time to load. So, I think you should...actually integrate it into an EMR system, so that the information can be accessed easily instead of through the ConnectingOntario interface. [P5]

Nonusers of the DHDR were unaware of how to register for access to the DHDR, and some participants in community settings highlighted the financial and administrative burden of gaining access to a provincial viewer. Others were unaware of the process of obtaining access, and some community physicians and pharmacists were unclear if they were eligible for access as they worked in independent practices.

Discussion

Principal Findings

To the best of our knowledge, this is the first evaluation of the DHDR repository in Ontario. Previous evaluations of similar national and international repositories have focused on the quality of data contained in the repository and how this affected clinical and process outcomes (eg, inappropriate prescribing, accuracy, and completeness of data elements, medication profiles, and compliance with BPMH documentation) [14,37,38]. Several evaluation studies found drug repositories to be incomplete and contain several discrepancies, thereby requiring use in conjunction with other sources to validate medication histories, and as such, drug repositories may be underused [14,39].

Our results highlight the untapped potential of the DHDR and provide insights into the key elements that are likely to drive future clinical value. Specifically, clinicians found the DHDR valuable in clinical situations where prompt communication with other clinicians or the pharmacy was not feasible, such as in the emergency department. A general lack of awareness was a significant barrier to realizing value at the system level, whereas a lack of comprehensive data was a barrier to consistent use. Moreover, the lack of integration into existing point-of-care systems was highlighted as a challenge to usage. Expanding the DHDR data set to include information on medication instructions and private insurance claims would enhance the clinical value and relative advantage of the DHDR.

The DHDR was most meaningful for clinicians who care for complex or older patients, such as geriatricians, underscoring the value of a comprehensive profile. Most of these patient populations are likely to be supported through publicly funded programs in which complete medication histories are captured within the DHDR. This suggests a potential use case in long-term care facilities, where polypharmacy is a common problem and the prevalence of ADEs is higher than that in other health care settings (ranging from 18% to 82%) [40]. Polypharmacy also increases the likelihood of hospital admission for patients living in long-term care [41], further emphasizing the potential untapped value of a centralized drug repository. Our sample had minimal representation from long-term care facilities, highlighting the need to further explore how to create value for clinicians working in this sector. The lack of medication information for most Ontario residents precluded meaningful use in most other settings. Only 20% of Ontario's 14.6 million residents are covered by public drug programs [22,42], limiting the DHDR's ability to deliver value for most of the population. Inclusion of all prescribed medications (whether dispensed) and the integration of community pharmacy records would support increased medication adherence and the reduction of inappropriate prescribing [43,44].

Despite a clear pathway to comprehensive data, a lack of awareness of the DHDR presents a fundamental barrier to achieving an impact at the population level. Although many participants perceived the data elements to be valuable, they were unaware that the repository existed or had limited knowledge on how to obtain access. This problem is not unique

to the DHDR [45] and highlights the need for active dissemination strategies, such as engaging local clinical champions and increasing awareness and education about the repository through professional organizations (eg, the Ontario Medical Association, the Registered Nurses Association of Ontario, or the Ontario College of Pharmacists), which may increase the uptake [46-48]. Cumbersome access pathways are not unique to the DHDR [45], suggesting that efforts to streamline onboarding, such as bundling it directly to licensure or credentialing processes rather than keeping it as an independent activity, would have positive downstream impacts beyond those realized through the DHDR. Integration into existing point-of-care systems is one such evidence-based strategy that would overcome a key barrier to adoption that plagues a range of digital technologies in health care [39,45,49,50] and has been successfully operationalized for a centralized drug repository [20].

Limitations

This formative evaluation aimed to capture responses from clinicians across multiple health care settings, as well as a broad sample of users and nonusers, to highlight areas for enhancements to the DHDR, which may increase meaningful clinical use. Consequently, our findings cannot speak to the realization of the impact on patient outcomes, experience, and system cost. Our findings are specific to sites with ≥20 registered users and may not reflect the experiences of providers in smaller centers (eg, rural communities). Although more female clinicians participated in our study, we feel this is comparable with the Ontario health care workforce, with over 70% being women. Moreover, we did not know how many individuals were sent the survey link, and as such, our findings may have selection bias because of voluntary self-selection.

As the survey was developed primarily by our project partners to inform decision-making, we sought to evaluate face validity and potential correlation across key domains (Multimedia Appendix 4). No statistical significance was observed, may be because of the limited number of active DHDR users, underscoring the need for further and more robust validity testing.

We were limited in our ability to assess the impact of the perceived shortcomings of the DHDR and associated clinical viewers. Teasing apart users' perceptions of the DHDR from their clinical viewer experience was beyond the scope of the

project; however, it is important to acknowledge. The DHDR was not co-designed by a representative sample of those using the repository, and only contained medication information for a subset of the Ontario population, which limited its clinical value. It is important to note that the DHDR was purposively implemented as an *incomplete* product by eHealth to facilitate the ability to gather data and insights to inform future development and expansion. Limited uptake suggests that future development efforts must consider marketing and implementation efforts to realize the impact. Finally, our objective was to understand health care provider perceptions, usage, and the clinical value of the DHDR to inform strategies to increase its utility and uptake. We determined that the uptake of the DHDR is a necessary precursor to its potential impact on care. Although patient experience is a critical next step in understanding whether and how such a tool can have an impact, it was beyond the scope of this project. An important evolution of this study is understanding how access to comprehensive medication data is valued by patients and how it can be used to enhance patient experience.

Conclusions

Findings from this formative evaluation suggest that the DHDR has untapped value as currently operationalized but that a pathway exists to align with clinician needs. Ensuring comprehensive clinical data and streamlined onboarding processes would facilitate meaningful use, whereas integration with existing point-of-care systems would further enhance efficiency and uptake. Access to a centralized repository that connects currently fragmented health care settings and provides comprehensive medication information at the point of care has the potential to improve efficiency and reduce medication errors if it aligns with the informational needs of clinical decision-making. Stakeholders involved in operationalizing and implementing such repositories (or similar information-sharing systems) should consider broad marketing and dissemination, user engagement, and integration into existing workflows as part of their overall strategy to realize value. Finally, user confidence in the DHDR can be improved by validating the information contained in the repository by comparing multiple sources of BPMH (ie, patient interviews vs pharmacy records vs DHDR records) and documenting the number of medication omissions, inappropriate prescribing, and discrepancies to validate DHDR. The knowledge that DHDR is accurate and complete will increase the uptake.

Authors' Contributions

CS designed the study, collected the data, interpreted and analyzed the data, and drafted the manuscript. MP collected the data and interpreted qualitative findings. MT, TJ, RSB, and LD designed the study and interpreted and analyzed the data. All authors edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Clinical Sensibility Questionnaire.
[[DOCX File , 53 KB](#) - formative_v6i3e27158_app1.docx]

Multimedia Appendix 2
Survey questions.
[[DOCX File , 73 KB](#) - formative_v6i3e27158_app2.docx]

Multimedia Appendix 3
Interview guide.
[[DOCX File , 21 KB](#) - formative_v6i3e27158_app3.docx]

Multimedia Appendix 4
Survey validation results for key experience domains.
[[DOCX File , 16 KB](#) - formative_v6i3e27158_app4.docx]

Multimedia Appendix 5
Comparison of demographic information in the survey for digital health drug repository (DHDR) users and nonusers (N=161).
[[DOCX File , 15 KB](#) - formative_v6i3e27158_app5.docx]

Multimedia Appendix 6
Frequency of digital health drug repository (DHDR) use among survey respondents (N=40).
[[DOCX File , 13 KB](#) - formative_v6i3e27158_app6.docx]

Multimedia Appendix 7
Mean scores and SDs of the digital health drug repository experience domains among survey respondents.
[[DOCX File , 16 KB](#) - formative_v6i3e27158_app7.docx]

Multimedia Appendix 8
Mean scores and SDs of the usefulness of current digital health drug repository (DHDR) data elements among DHDR users in the survey.
[[DOCX File , 17 KB](#) - formative_v6i3e27158_app8.docx]

Multimedia Appendix 9
Rankings for new digital health drug repository (DHDR) data elements among survey respondents.
[[DOCX File , 14 KB](#) - formative_v6i3e27158_app9.docx]

Multimedia Appendix 10
Overall perceptions of the value and impact of the digital health drug repository (DHDR) among female and male DHDR users (N=40).
[[DOCX File , 15 KB](#) - formative_v6i3e27158_app10.docx]

Multimedia Appendix 11
Outstanding quotes.
[[DOCX File , 19 KB](#) - formative_v6i3e27158_app11.docx]

References

1. Rotermann M, Sanmartin C, Hennessy D, Arthur M. Prescription medication use by Canadians aged 6 to 79. Health Rep 2014;25(6):3-9 [FREE Full text] [Medline: 24941315]
2. Cornish PL, Knowles SR, Marchesano R, Tam V, Shadowitz S, Juurlink DN, et al. Unintended medication discrepancies at the time of hospital admission. Arch Intern Med 2005;165(4):424-429. [doi: 10.1001/archinte.165.4.424] [Medline: 15738372]

3. Forster AJ, Murff HJ, Peterson JF, Gandhi TK, Bates DW. The incidence and severity of adverse events affecting patients after discharge from the hospital. Ann Intern Med 2003;138(3):161-167. [doi: 10.7326/0003-4819-138-3-200302040-00007] [Medline: 12558354]

4. Laatikainen O, Miettunen J, Sneck S, Lehtiniemi H, Tenhunen O, Turpeinen M. The prevalence of medication-related adverse events in inpatients-a systematic review and meta-analysis. Eur J Clin Pharmacol 2017;73(12):1539-1549. [doi: 10.1007/s00228-017-2330-3] [Medline: 28871436]

5. Gurwitz JH, Field TS, Harrold LR, Rothschild J, Debellis K, Seger AC, et al. Incidence and preventability of adverse drug events among older persons in the ambulatory setting. JAMA 2003;289(9):1107-1116. [doi: 10.1001/jama.289.9.1107] [Medline: 12622580]

6. Annual demographic estimates: Canada, provinces and territories, 2020. Statistics Canada. 2020. URL: https://www150.statcan.gc.ca/n1/pub/91-215-x/91-215-x2020001-eng.htm [accessed 2018-11-20]

7. United States Census. U.S. population (2020). URL: https://www.census.gov/popclock/ [accessed 2018-11-20]

8. Assiri GA, Shebl NA, Mahmoud MA, Aloudah N, Grant E, Aljadhey H, et al. What is the epidemiology of medication errors, error-related adverse events and risk factors for errors in adults managed in community care contexts? A systematic review of the international literature. BMJ Open 2018;8(5):e019101 [FREE Full text] [doi: 10.1136/bmjopen-2017-019101] [Medline: 29730617]

9. Etchells E, Mittmann N, Koo M, Baker M, Krahn M, Shojania K, et al. The economics of patient safety in acute care: technical report. Canadian Patient Safety Institute. 2012. URL: https://www.patientsafetyinstitute.ca/en/toolsResources/Research/commissionedResearch/EconomicsofPatientSafety/Documents/Economics%20of%20Patient%20Safety%20-%20Acute%20Care%20-%20Final%20Report.pdf [accessed 2018-11-20]

10. Medication reconciliation (MedRec). Canada Institute for Safe Medication Practices Canada. URL: https://www.ismp-canada.org/medrec/ [accessed 2018-11-20]

11. Wang H, Meng L, Song J, Yang J, Li J, Qiu F. Electronic medication reconciliation in hospitals: a systematic review and meta-analysis. Eur J Hosp Pharm 2018;25(5):245-250 [FREE Full text] [doi: 10.1136/ejhpharm-2017-001441] [Medline: 31157034]

12. McNab D, Bowie P, Ross A, MacWalter G, Ryan M, Morrison J. Systematic review and meta-analysis of the effectiveness of pharmacist-led medication reconciliation in the community after hospital discharge. BMJ Qual Saf 2018;27(4):308-320 [FREE Full text] [doi: 10.1136/bmjqs-2017-007087] [Medline: 29248878]

13. Medication reconciliation: a learning guide. Queens's Univeristy. URL: https://meds.queensu.ca/central/assets/modules/mr/1.html [accessed 2018-11-20]

14. Kalb K, Shalansky S, Legal M, Khan N, Ma I, Hunte G. Unintended medication discrepancies associated with reliance on prescription databases for medication reconciliation on admission to a general medical ward. Can J Hosp Pharm 2009;62(4):284-289 [FREE Full text] [doi: 10.4212/cjhp.v62i4.809] [Medline: 22478906]

15. Armstrong S. Patient access to health records: striving for the Swedish ideal. BMJ 2017;357:j2069. [doi: 10.1136/bmj.j2069] [Medline: 28465316]

16. Australia's medication cloud powered by eRx script exchange. MedView. URL: https://www.medview.com.au/ [accessed 2018-11-20]

17. NHI MediCloud system. National Health Insurance Administration Ministry of Health and Welfare. URL: https://www.nhi.gov.tw/english/Content_List.aspx?n=02BA04454AED80E0&topn=BCB2B0D2433F6491 [accessed 2018-11-20]

18. Dormuth CR, Miller TA, Huang A, Mamdani MM, Juurlink DN, Canadian Drug Safety and Effectiveness Research Network. Effect of a centralized prescription network on inappropriate prescriptions for opioid analgesics and benzodiazepines. CMAJ 2012;184(16):E852-E856 [FREE Full text] [doi: 10.1503/cmaj.120465] [Medline: 22949563]

19. Dahri K, Shalansky SJ, Jang L, Jung L, Ignaszewski AP, Clark C. Accuracy of a provincial prescription database for assessing medication adherence in heart failure patients. Ann Pharmacother 2008;42(3):361-367. [doi: 10.1345/aph.1K385] [Medline: 18303147]

20. Huang SK, Wang PJ, Tseng WF, Syu FK, Lee MC, Shih RL, et al. NHI-PharmaCloud in Taiwan--a preliminary evaluation using the RE-AIM framework and lessons learned. Int J Med Inform 2015;84(10):817-825. [doi: 10.1016/j.ijmedinf.2015.06.001] [Medline: 26113462]

21. Johnson JA, Bootman JL. Drug-related morbidity and mortality. A cost-of-illness model. Arch Intern Med 1995;155(18):1949-1956. [Medline: 7575048]

22. Ontario public drug programs. Ministry of Health Ontario. URL: http://www.health.gov.on.ca/en/pro/programs/drugs/funded_drug/funded_drug.aspx [accessed 2018-11-20]

23. Narcotics monitoring system. Ministry of Health Ontario. URL: http://www.health.gov.on.ca/en/pro/programs/drugs/ons/monitoring_system.aspx [accessed 2018-11-20]

24. Auspitz M, Cleghorn MC, Azin A, Sockalingam S, Quereshy FA, Okrainec A, et al. Knowledge and perception of bariatric surgery among primary care physicians: a survey of family doctors in Ontario. Obes Surg 2016;26(9):2022-2028. [doi: 10.1007/s11695-016-2055-x] [Medline: 26780362]

XSL•FO
RenderX

25. Malhotra J, Wong E, Thind A. Canadian family physician job satisfaction - is it changing in an evolving practice environment? An analysis of the 2013 National Physician Survey database. BMC Fam Pract 2018;19(1):100 [FREE Full text] [doi: 10.1186/s12875-018-0786-6] [Medline: 29935531]

26. Burns KE, Duffett M, Kho ME, Meade MO, Adhikari NK, Sinuff T, ACCADEMY Group. A guide for the design and conduct of self-administered surveys of clinicians. CMAJ 2008;179(3):245-252 [FREE Full text] [doi: 10.1503/cmaj.080372] [Medline: 18663204]

27. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]

28. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]

29. R project for statistical computing. The R Foundation. URL: http://www.r-project.org/#! [accessed 2018-11-20]

30. Bradshaw C, Atkinson S, Doody O. Employing a qualitative description approach in health care research. Glob Qual Nurs Res 2017;4:2333393617742282 [FREE Full text] [doi: 10.1177/2333393617742282] [Medline: 29204457]

31. Caelli K, Ray L, Mill J. 'Clear as Mud': toward greater clarity in generic qualitative research. Int J Qual Methods 2003;2(2):1-13. [doi: 10.1177/160940690300200201]

32. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]

33. A place to organize, store and analyze your data. NVivo. 2020. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2018-11-20]

34. DeVon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ, et al. A psychometric toolbox for testing validity and reliability. J Nurs Scholarsh 2007;39(2):155-164. [doi: 10.1111/j.1547-5069.2007.00161.x] [Medline: 17535316]

35. Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. 5th edition. Oxford, UK: Oxford University Press; 2015.

36. Daniel WW, Cross CL. Biostatistics: a foundation for analysis in health sciences. 10th edition. Hoboken, NJ: John Wiley & Sons; 2019.

37. Karkov LL, Schytte-Hansen S, Haugbølle LS. Discrepancies between sources providing the medication histories of acutely hospitalised patients. Pharm World Sci 2010;32(4):449-454. [doi: 10.1007/s11096-010-9390-0] [Medline: 20446111]

38. Motulsky A, Weir DL, Couture I, Sicotte C, Gagnon MP, Buckeridge DL, et al. Usage and accuracy of medication data from nationwide health information exchange in Quebec, Canada. J Am Med Inform Assoc 2018;25(6):722-729 [FREE Full text] [doi: 10.1093/jamia/ocy015] [Medline: 29590350]

39. Scott IA, Pillans PI, Barras M, Morris C. Using EMR-enabled computerized decision support systems to reduce prescribing of potentially inappropriate medications: a narrative review. Ther Adv Drug Saf 2018;9(9):559-573 [FREE Full text] [doi: 10.1177/2042098618784809] [Medline: 30181862]

40. Storms H, Marquet K, Aertgeerts B, Claes N. Prevalence of inappropriate medication use in residential long-term care facilities for the elderly: a systematic review. Eur J Gen Pract 2017;23(1):69-77 [FREE Full text] [doi: 10.1080/13814788.2017.1288211] [Medline: 28271916]

41. Afonso-Argilés FJ, Meyer G, Stephan A, Comas M, Wübker A, Leino-Kilpi H, RightTimePlaceCare Consortium. Emergency department and hospital admissions among people with dementia living at home or in nursing homes: results of the European RightTimePlaceCare project on their frequency, associated factors and costs. BMC Geriatr 2020;20(1):453 [FREE Full text] [doi: 10.1186/s12877-020-01835-x] [Medline: 33153444]

42. Population estimates, quarterly: table 17-10-0009-01. Statisitics Canada. 2021. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000901 [accessed 2020-06-26]

43. Wright DJ, Twigg MJ. Community pharmacy: an untapped patient data resource. Integr Pharm Res Pract 2016;5:19-25 [FREE Full text] [doi: 10.2147/IPRP.S83261] [Medline: 29354535]

44. Rankin A, Cadogan CA, Patterson SM, Kerse N, Cardwell CR, Bradley MC, et al. Interventions to improve the appropriate use of polypharmacy for older people. Cochrane Database Syst Rev 2018;9(9):CD008165 [FREE Full text] [doi: 10.1002/14651858.CD008165.pub4] [Medline: 30175841]

45. Wickerson L, Fujioka JK, Kishimoto V, Jamieson T, Fine B, Bhatia RS, et al. Utility and perceived value of a provincial digital diagnostic imaging repository: a multimethods study. JMIR Form Res 2020;4(7):e17220 [FREE Full text] [doi: 10.2196/17220] [Medline: 32459644]

46. Hempel S, O'Hanlon C, Lim YW, Danz M, Larkin J, Rubenstein L. Spread tools: a systematic review of components, uptake, and effectiveness of quality improvement toolkits. Implement Sci 2019;14(1):83 [FREE Full text] [doi: 10.1186/s13012-019-0929-8] [Medline: 31426825]

47. Gagnon ML. Moving knowledge to action through dissemination and exchange. J Clin Epidemiol 2011;64(1):25-31. [doi: 10.1016/j.jclinepi.2009.08.013] [Medline: 20004552]

48.   Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PC, et al. Why don't physicians follow clinical practice
      guidelines? A framework for improvement. JAMA 1999;282(15):1458-1465. [doi: 10.1001/jama.282.15.1458] [Medline:
      10535437]
49.   Keasberry J, Scott IA, Sullivan C, Staib A, Ashby R. Going digital: a narrative overview of the clinical and organisational
      impacts of eHealth technologies in hospital practice. Aust Health Rev 2017;41(6):646-664. [doi: 10.1071/AH16233]
      [Medline: 28063462]
50.   Marceglia S, D'Antrassi P, Prenassi M, Rossi L, Barbieri S. Point of care research: integrating patient-generated data into
      electronic health records for clinical trials. AMIA Annu Symp Proc 2018;2017:1262-1271 [FREE Full text] [Medline:
      29854195]

Abbreviations

ADE: adverse drug event
BPMH: best possible medication history
DHDR: Digital Health Drug Repository
LRA: Local Registration Authority

Original Paper

A Free Open-Source Bayesian Vancomycin Dosing App for Adults: Design and Evaluation Study

Thomas Oommen[1], PharmD; Anirudh Thommandram[2], MASc; Adam Palanica[2], PhD; Yan Fossat[2], CAP

[1]Pharmacy, Lakeridge Health, Oshawa, ON, Canada

[2]Klick Applied Sciences, Klick Health, Klick Inc, Toronto, ON, Canada

Corresponding Author:
Thomas Oommen, PharmD
Pharmacy
Lakeridge Health
1 Hospital Court
Oshawa, ON, L1G 2B9
Canada
Phone: 1 9055768711
Email: toommen@lh.ca

Abstract

Background: It has been suggested that Bayesian dosing apps can assist in the therapeutic drug monitoring of patients receiving vancomycin. Unfortunately, Bayesian dosing tools are often unaffordable to resource-limited hospitals. Our aim was to improve vancomycin dosing in adults. We created a free and open-source dose adjustment app, VancoCalc, which uses Bayesian inference to aid clinicians in dosing and monitoring of vancomycin.

Objective: The aim of this paper is to describe the design, development, usability, and evaluation of a free open-source Bayesian vancomycin dosing app, VancoCalc.

Methods: The app build and model fitting process were described. Previously published pharmacokinetic models were used as priors. The ability of the app to predict vancomycin concentrations was performed using a small data set comprising of 52 patients, aged 18 years and over, who received at least 1 dose of intravenous vancomycin and had at least 2 vancomycin concentrations drawn between July 2018 and January 2021 at Lakeridge Health Corporation Ontario, Canada. With these estimated and actual concentrations, median prediction error (bias), median absolute error (accuracy), and root mean square error (precision) were calculated to evaluate the accuracy of the Bayesian estimated pharmacokinetic parameters.

Results: A total of 52 unique patients' initial vancomycin concentrations were used to predict subsequent concentration; 104 total vancomycin concentrations were assessed. The median prediction error was –0.600 ug/mL (IQR –3.06, 2.95), the median absolute error was 3.05 ug/mL (IQR 1.44, 4.50), and the root mean square error was 5.34.

Conclusions: We described a free, open-source Bayesian vancomycin dosing calculator based on revisions of currently available calculators. Based on this small retrospective preliminary sample of patients, the app offers reasonable accuracy and bias, which may be used in everyday practice. By offering this free, open-source app, further prospective validation could be implemented in the near future.

Introduction

Vancomycin is an intravenous (IV) drug that has been used for over 60 years in the treatment of Gram-positive bacterial infections. Vancomycin has a narrow therapeutic window; thus, therapeutic drug monitoring (TDM) of plasma concentrations is necessary to measure efficacy and to avoid nephrotoxicity. Traditionally, vancomycin dosing consists of giving a patient-weight–based dose, and then checking a trough concentration at an approximate steady state concentration, targeting a range between 10 and 20 mg/L [1]. Dose adjustments

are often made using simple heuristics, clinical intuition, and judgement based on patient factors.

In 2020, the Infectious Diseases Society of America updated the dosing guidelines of vancomycin for severe methicillin-resistant *Staphylococcus aureus* infections. They suggest that vancomycin monitoring via an area under the curve (AUC) approach is the safest and most effective approach to TDM [1]. AUC measurement can be calculated by taking blood concentrations and performing pharmacokinetic (PK) calculations to derive an AUC estimate. This is measured by either the Sawchuk-Zaske method (ie, based on individually calculated PK parameters) or using a Bayesian inference approach (ie, a method of statistical inference using the Bayes theorem to update the probability as more data become available) [2]. According to the Infectious Diseases Society of America, Bayesian-based tools are the preferred approach to vancomycin TDM [1]. Controversy still exists regarding the adoption of AUC-only monitoring of vancomycin [3-5]. Some suggest that adopting this approach may be premature [3], and AUC dosing if carried out by 2-level approach may increase the patient burden by increasing blood draws. There is still some uncertainty on the ideal target AUC or trough level [4].

Calls for individualized Bayesian dosing tools have existed for years [6-10]; however, Bayesian TDM has failed to garner widespread adoption for many reasons [7]. The Bayesian TDM approach can require paid computer software programs. These programs typically require multiyear licensing agreements and can be unaffordable to hospital systems. These software packages often use publicly available PK models [11-14] and Bayesian algorithms to derive dosing suggestions. Most apps do not make their code public due to proprietary coding and financial interest. However, it may be assumed that the PK models and Bayesian methods are not proprietary. These Bayesian TDM apps make use of these PK models as a priori to assist in predicting individual patient PK parameters. The

apps conduct mathematical calculations seamlessly without requiring the user to have a knowledge of statistics. The Bayesian inference algorithms were once limited by computing power; however, with advances in technology, they can be performed with any basic software browser via HTML and JavaScript.

As a solution to these paid software programs, we developed a proof-of-concept, free online Bayesian vancomycin dosing app, VancoCalc [15]. VancoCalc allows vancomycin TDM using trough or AUC-based targets. The app uses published PK models and user-inputted patient data as Bayesian priors to estimate vancomycin concentrations and AUC. The app requires no statistical training and is aimed to be user-friendly to assist in the implementation of Bayesian inference in vancomycin TDM. No user data are saved as all calculations are computed locally on the user's device.

Inspiration for our new calculator was fostered from anecdotal user experience with DosOpt [16], ClinCalc [17], and BestDose [18], in hopes of making a more user-friendly app. VancoCalc adds the ability to customize target trough or AUC, permitting 2 different dose or time intervals (plus loading dose) relating to the vancomycin concentration, and the ability to explore various dosing regimens through visualization of the dose and time curve. These updates assist the clinician in assessing the impact a dosing regimen may have (Figure 1). After inputting "Bayesian Vancomycin Calculator" in a search engine, it may appear that there are many Bayesian dosing apps. However, many of these calculators use 2-level trapezoidal calculations and are not truly using Bayesian approaches.

The app's ability to predict a patient's vancomycin concentration (pharmacokinetic parameters) was evaluated using a data set composed of adult patients who received vancomycin at Lakeridge Health. This research aims to describe the development, design, and evaluation of a Bayesian vancomycin dosing app.

Figure 1. Example of the app's graphical user interface. AUC: area under the curve; MIC: minimum inhibitory concentration; PK: pharmacokinetic.
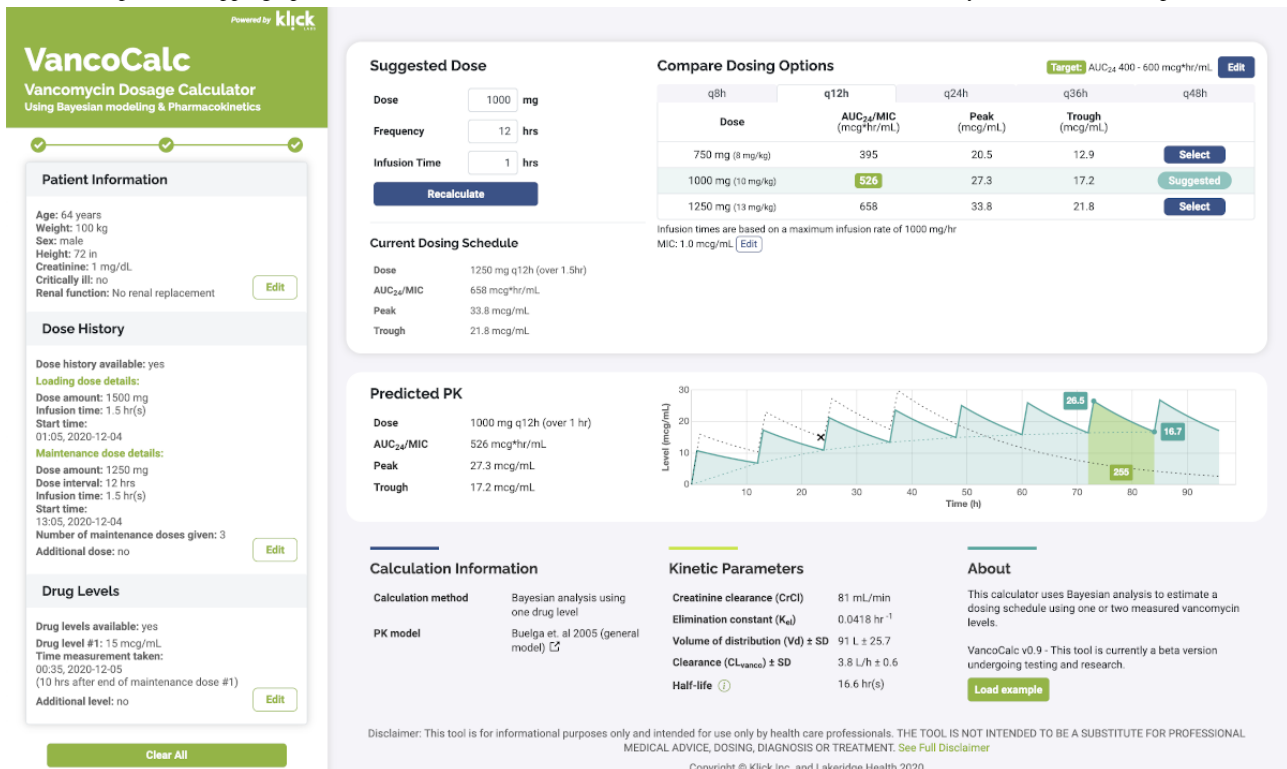
Methods

Development of VancoCalc App

We implemented the app as a static web app using HTML5, JavaScript, CSS, and several open-source projects. The user interface was developed using the jQuery [19], Bootstrap [20]. And Chart.js [21] frameworks. The app uses a custom library for applying Bayesian algorithms that is based on bayes.js [22] and the simple statistics [23] libraries. Markov chain Monte Carlo sampling and model fitting are carried out in the user's browser with no data processing needed from a server or the cloud.

App Design Bayesian Pharmacokinetic Modelling and Fitting

Bayesian dosing strategies employ population models that relate PK parameters, including volume of distribution (VD), clearance (Cl), and creatinine clearance to patient data, including age, weight, gender, serum creatinine, vancomycin dose value, and number of doses given. The Bayesian approach involves the notion of incorporating both a population PK model and measured drug concentrations from the patient to better estimate PK parameters for the individual.

The app selects a population model based on the inputted patient data. If a patient is critically ill and has a BMI of over 30 kg/m^2 and body weight of over 100 kg, the Masich et al [14] model is used; if a patient is critically ill but is not meeting the obese criteria, the Roberts et al [13] model is used; if not critically ill but has a BMI of over 40 kg/m^2 and body weight of over 120 kg, then the Adane et al [11] model is used; if not critically ill and BMI is less than 40 kg/m^2, the Buelga et al [12] general population model is used. The critically ill model parameter is

possible to override. Estimates of the PK parameters (VD and Cl) from the population model, adjusted for patient characteristics, form the starting point of the fitting process. The starting point begins with the mean VD and Cl of the PK model selected. The values are sampled from the parameter space, and first-order PK equations are used to calculate an expected serum concentration, which is compared to the measured concentration. This sampling and traversing of the parameter space are carried out iteratively while taking into consideration the variability of the population parameters and the variability of the serum concentration measurement. By doing this, the modelling process estimates the PK parameters that will be most consistent with drug concentrations predicted by both the population model and the measured drug concentrations.

Complex dose history can be entered involving multiple doses of varying amounts, frequencies, and schedules. Concentrations are not restricted to steady state and can be entered for any infusion and are not limited to trough levels. The calculator's inputs and outputs are shown on the same page, side-by-side, and any modifications to the input are immediately reflected in the calculated output dynamically. The design provides a suggested dose and alternatives, but also encourages exploration and investigation for the patient cases that require it.

App Evaluation

A retrospective observational data set was used to evaluate the app. Information was gathered via electronic chart review (Meditech) and SQL query of Antimicrobial Stewardship data repository.

The data were collected between July 2018 and January 2021 at Lakeridge Health Corporation sites in Oshawa, Bowmanville, or Port Perry, Ontario, Canada. This is a community hospital

system with approximately 800 beds located in the Durham region of Ontario. Inclusion criteria were inpatients, aged 18 years and over, who received at least 1 dose of IV vancomycin and had at least 2 vancomycin concentrations drawn in relation to the vancomycin dose. Patient's sex, age, height, weight, serum creatinine at time of vancomycin, ward, vancomycin dosing history, and vancomycin concentrations were collected. Exclusion criteria were patients who were receiving hemodialysis or continuous renal replacement while on vancomycin, or missing any data as stated in the inclusion criteria.

The first single known vancomycin concentration and matching patient variables were entered into the app. This allowed the app to estimate the patient's individual PK parameters. Subsequent vancomycin concentrations (if available) were not inputted as the app currently only allows 2 vancomycin concentrations to be inputted. This produced a vancomycin plasma concentration time plot where the estimated second vancomycin concentration was compared to the actual concentration.

Ethics Approval

The Research Ethics Board (REB) at Lakeridge Health approved the study (Approval 2020-13) in September 2020, Oshawa, Ontario.

Statistical Analysis and Predictive Performance

Performance of Bayesian prediction of the second vancomycin concentration was evaluated with the median prediction error (MedPE), median absolute error (MedAE), and root mean squared error (RMSE). MedPE, MedAE, and RMSE were calculated according to Equations 1, 2, and 3, and were derived from the study by Sheiner and Beal [24]. MedPE was calculated as an index of bias, MedAE as an index of accuracy, and RMSE as an index of precision. Median was used as the data displayed a nonnormal distribution. All analyses were performed using R (The R Foundation) [25].


Funding

Results

A total of 93 patient charts were reviewed; 41 charts were excluded for analysis. The main reason for exclusion was the patients being on renal replacement therapy. Twenty-five cases were excluded as they were on renal replacement therapy; 2 cases were excluded due to being pediatric; 2 patients were excluded for model fit; and 12 patients had to be excluded where we were unable to reliably input the patient dose parameters (staggered frequency of dose administration or multiple dosing changes; subsequent vancomycin concentration did not correspond with inputted dose). A total of 104 vancomycin concentrations were assessed. The remaining 52 patients were included in the final analysis. This included 24 (46.2%) female and 28 (53.8%) male participants; age range 24-90 years, median age 63 (SD 14) years; median weight 81.2 kg, weight range 43.6-131.7 (SD 20.8) kg; median BMI 30.7 kg/m^2, BMI range 16-54.1 (SD 8.6) kg/m^2. Moreover, 15 (28.8% of the 52 patients were admitted to the intensive care unit.

The median measured concentration was 18.05 (SD 6.83) ug/mL versus the median of the predicted concentration 18.49 (SD 6.40) ug/mL. Using Bayesian estimation, MedPE (bias) was –0.600 ug/mL (IQR –3.06, 2.95), the MedAE (accuracy) was 3.05 ug/mL (IQR 1.44, 4.50), and the RMSE (precision) was 5.34. These results were derived from the following a priori models: Adane et al [11] (n=1); Buelga et al [12] (n=36); Masich et al [14] (n=4); and Roberts et al [13] (n=11; Figure 2). Of note, there were 2 outliers noted in the analysis. One patient was morbidly obese, and the other critically ill. These 2 patients were fit to the Adane et al [11] and Roberts et al [13] a priori models. These outliers were included in the analysis as we could not identify a specific reason for these prediction errors other than their underlying conditions. Performing the analysis excluding these outliers improves median accuracy and precision, with the median bias being largely unchanged. Excluding these outliers, the MedPE was –0.6 ug/mL (IQR –2.9, 2.75), MedAE was 2.95 ug/mL (IQR 1.40, 4.16), and RMSE was 3.86. As other publications have previously reported mean prediction error and mean absolute prediction error [26], we have also included these results. Mean prediction error was –3.13 ug/mL, and mean absolute prediction error was 3.69 ug/mL.

Table 1 displays the bias, accuracy, and RMSE of the individual models. Interpretation of these values is difficult due to the small sample size of some of the patients included in the models.

Figure 2. Observed versus predicted vancomycin concentration and an a priori model selected (N=52 from the following models: Adane, n=1; Buelga, n=36; Masich, n=4; and Roberts, n=11).
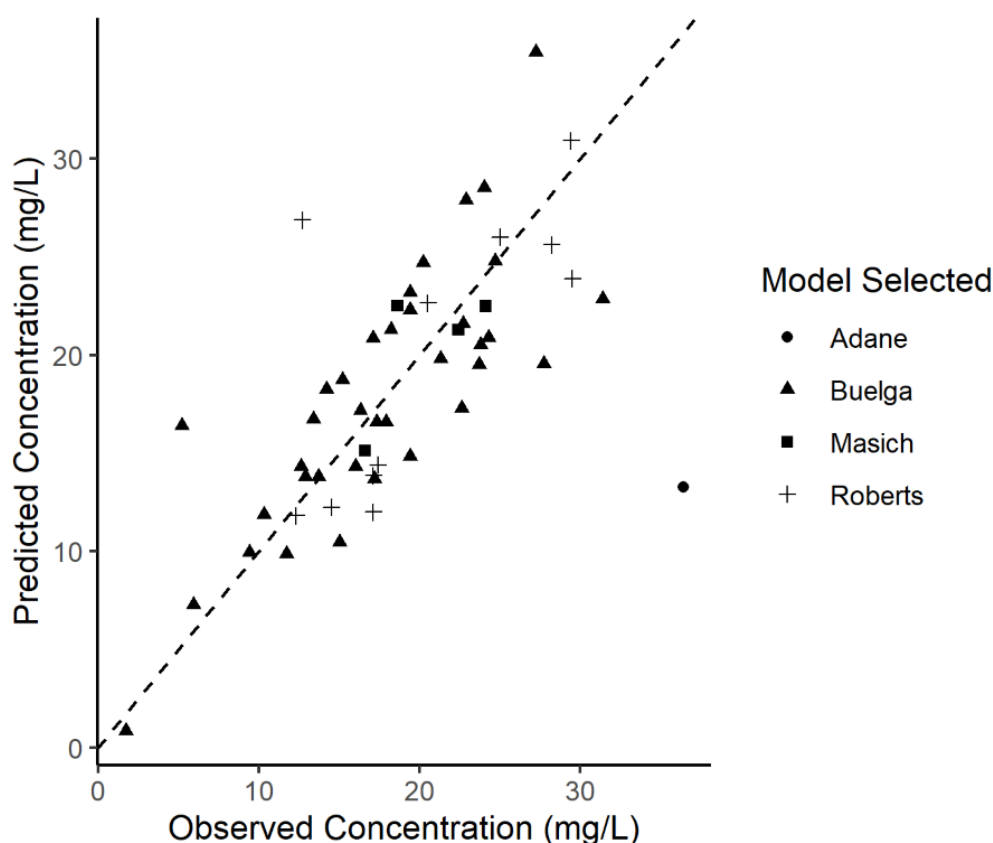
Table 1. App performance of individual pharmacokinetic models.

PK[a] model	Unique patients, n (%)	Median prediction error (mcg/mL)	Median absolute error (mcg/mL)	Root mean square error
Adane et al [11]	1 (1.9)	−23.14	23.14	23.14
Buelga et al [12]	36 (69.2)	0.34	3.33	4.18
Masich et al [14]	4 (7.7)	−1.29	1.54	2.30
Roberts et al [13]	11 (21.2)	−2.26	2.59	5.20

[a]PK: pharmacokinetic.

Discussion

Principal Results

We created a free and open-source Bayesian vancomycin dosing app. No statistical or technical knowledge of Bayesian methods are required to use this app. We focused on making this app easy to use for clinicians, with an ability to explore treatment adjustments.

Vancomycin concentrations in second blood samples were predicted by Bayesian analysis and were compared with measured concentrations to assess the app and PK model accuracy of predictions. Our data represent a small sample size of patients in the Durham region of Ontario. Our total patient sample of 52 patients is small. Hiraki et al [26], Turner et al [27], and Nunn et al [28] all used similar sample sizes in their analysis of Bayesian vancomycin dosing software and PK models.

Comparisons With Prior Work

We reviewed the findings of Hiraki et al [26], Turner et al [27], and Nunn et al [28]. We chose to use the methods described by Sheiner and Beal [24] to report our predictive accuracy. Hiraki et al [26] did the same; however, other publications chose alternate approaches to reporting predictive accuracy [27,28]. When comparing our prediction accuracy versus those reported by Hiraki et al [26], we showed similar bias and accuracy. However, our overall results were not as precise as the results shared by Hiraki et al [26] (RMSE 5.34 vs 1.74). We attribute this to the 2 outliers within our data set, and as such, these patients may not be ideal candidates for the a priori model selected. Additionally, our data set included a more heterogeneous population of both critically ill and on the general medicine wards, who were being treated with vancomycin for multiple reasons. The population studied by Hiraki et al [26] was limited to hospitalized patients receiving treatment for methicillin-resistant *S. aureus* infections whose renal function

was very stable. This could also explain the difference in precision.

Interpretation and performance of the individual PK model accuracy and bias in our paper is limited due to the small sample but are included for review (Table 1). Only 1 patient in the validation data set was fit to the Adane et al [11] model, and 4 patients to the Masich et al [14] model. The predicted result of the single patient fit to the Adane et at [11] model and the one fit to the Roberts et al [13] model produced predicted concentrations that were outliers. The Adane et al [11] model is reserved for patients with a BMI greater than 40 kg/m^2 and body weight over 120 kg, while the Robers et al [13] model is reserved for the critically ill. Although there were only 2 data points, this highlights the difficulty in dosing these patients. Based on the small number of patients we enrolled during the time frame of this study, it is unlikely that we will be able to obtain a sufficient sample size to evaluate these models individually.

Limitations

We acknowledge that there are limitations of this app. First, it remains to be determined which a priori PK model is optimal for dosing, since many published vancomycin PK models exist, and choosing the appropriate PK model remains a challenge. It may be unlikely that the wide individual patient PK variability can be captured by a single a priori model. Therefore, it is of utmost importance that the patient be appropriate for the a priori model selected. We chose a one-compartment pharmacokinetic model with individual models for obesity and critical illness. During the development of the app, it appears that a more simplified broadly supported two-compartment model by Goti et al [29] may be more appropriate [30,31].

As trough-based vancomycin dosing was performed at Lakeridge, we are unable evaluate the app's AUC predictions. However, providing this app free of charge for all users allows the potential for further evaluation of this functionality.

The app is not designed to be used in patients receiving renal replacement therapy. Many patients were excluded from the evaluation as they were receiving vancomycin while on renal replacement therapy.

This app was not intended to replace or override the judgement of a clinician. The app relies on accurate user inputs. There are a few checks in place to ensure appropriate user input and model selection (eg, hard limits on input parameters and goodness of fit with the model), but we acknowledge that all possible scenarios or edge cases cannot be addressed. In the development of the app, we attempted to balance user experience and ease of use versus added complexity and customization of the app and a priori PK models. As suggested earlier, a potential solution to this juxtaposition is a more generally applied model similar to that presented by Goti et al [29], or, if suggested, a more customized user-selected model approach or "expert mode."

Future Directions

Bayesian dosing of vancomycin still in its infancy. Vancocalc is available with no restriction. The creation of this free, open-source Bayesian vancomycin dosing app permits this dosing approach to be further explored.

The value of free, open-source software is unquantifiable. Software such as Linux and Python have positively impacted many. We are hopeful that the creation of this app can also have a positive impact and add to the body of work that has already been conducted [16,17].

With increased use and validation of Bayesian dosing apps, it is possible that improved PK models for specific patient populations will be developed. Additionally, larger adoption may eventually permit improved algorithms incorporating other techniques such as machine learning. By open sourcing the app, it can be modified, updated, and improved upon or used as a framework for others to build upon.

Our goal is to share this proof-of-concept app to allow greater awareness and democratization of the Bayesian monitoring approach. The pharmacokinetics of vancomycin are complex. With greater use and awareness of the app and experience from field experts, this will lead to further collaboration and improvement of the app for functionality and vancomycin TDM.

Conclusions

In this paper, we presented a vancomycin dosing app, VancoCalc, including a small evaluation using a real-world retrospective data set. This app leverages previously published research on Bayesian inference calculators. By offering this free, open-source app, further prospective validation could be implemented in the near future. We encourage exploration of the app and collaboration or suggestions for improvement.

Authors' Contributions

TO executed the study, analyzed the data, and contributed to writing the manuscript. AT developed the data modelling, developed the VancoCalc website, and helped analyze the data. AP contributed to literature review and writing the manuscript. YF helped design the study, assisted in data interpretation, and contributed to the writing and editing of the manuscript. All authors contributed to the final review and editing and have approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Rybak MJ, Le J, Lodise TP, Levine DP, Bradley JS, Liu C, et al. Therapeutic monitoring of vancomycin for serious methicillin-resistant Staphylococcus aureus infections: A revised consensus guideline and review by the American Society of Health-System Pharmacists, the Infectious Diseases Society of America, the Pediatric Infectious Diseases Society, and the Society of Infectious Diseases Pharmacists. Am J Health Syst Pharm 2020 May 19;77(11):835-864. [doi: 10.1093/ajhp/zxaa036] [Medline: 32191793]

2. Neely MN, Youn G, Jones B, Jelliffe RW, Drusano GL, Rodvold KA, et al. Are vancomycin trough concentrations adequate for optimal dosing? Antimicrob Agents Chemother 2014;58(1):309-316 [FREE Full text] [doi: 10.1128/AAC.01653-13] [Medline: 24165176]

3. Dalton BR, Rajakumar I, Langevin A, Ondro C, Sabuda D, Griener TP, et al. Vancomycin area under the curve to minimum inhibitory concentration ratio predicting clinical outcome: a systematic review and meta-analysis with pooled sensitivity and specificity. Clin Microbiol Infect 2020 Apr;26(4):436-446 [FREE Full text] [doi: 10.1016/j.cmi.2019.10.029] [Medline: 31756451]

4. The Canadian Journal of Hospital Pharmacy. Should therapeutic monitoring of vancomycin based on area under the curve become standard practice for patients with confirmed or suspected methicillin-resistant infection? Can J Hosp Pharm 2020;73(3):232-237 [FREE Full text] [Medline: 32616951]

5. Jorgensen SCJ, Stewart JJ, Dalton B. Vancomycin AUC-guided therapeutic drug monitoring to reduce nephrotoxicity: are we overlooking a simpler solution? Comment on: Oda K, Jono H, Nosaka K, Saito H. Reduced nephrotoxicity in vancomycin therapeutic drug monitoring guided by area under the concentration-time curve against trough 15-20 µg/mL concentration. Int J Antimicrob Agents 2020 Nov;56(5):106150. [doi: 10.1016/j.ijantimicag.2020.106150] [Medline: 32882431]

6. Donagher J, Martin JH, Barras MA. Individualised medicine: why we need Bayesian dosing. Intern Med J 2017 May;47(5):593-600. [doi: 10.1111/imj.13412] [Medline: 28503880]

7. Kufel WD, Seabury RW, Mogle BT, Beccari MV, Probst LA, Steele JM. Readiness to implement vancomycin monitoring based on area under the concentration-time curve: A cross-sectional survey of a national health consortium. Am J Health Syst Pharm 2019 Jun 03;76(12):889-894. [doi: 10.1093/ajhp/zxz070] [Medline: 31063582]

8. Neely MN, Kato L, Youn G, Kraler L, Bayard D, van Guilder M, et al. Prospective trial on the use of trough concentration versus area under the curve to determine therapeutic vancomycin dosing. Antimicrob Agents Chemother 2018 Feb;62(2):1-12 [FREE Full text] [doi: 10.1128/AAC.02042-17] [Medline: 29203493]

9. Uaamnuichai M, Day RB, Brater DC. Bayesian and least-squares methods for vancomycin dosing. Am J Med Sci 1987 Aug;294(2):100-104. [doi: 10.1097/00000441-198708000-00008] [Medline: 3631119]

10. Avent ML, Rogers BA. Optimising antimicrobial therapy through the use of Bayesian dosing programs. Int J Clin Pharm 2019 Oct;41(5):1121-1130. [doi: 10.1007/s11096-019-00886-4] [Medline: 31392582]

11. Adane ED, Herald M, Koura F. Pharmacokinetics of vancomycin in extremely obese patients with suspected or confirmed Staphylococcus aureus infections. Pharmacotherapy 2015 Feb;35(2):127-139. [doi: 10.1002/phar.1531] [Medline: 25644478]

12. Buelga DS, del Mar Fernandez de Gatta M, Herrera EV, Dominguez-Gil A, García MJ. Population pharmacokinetic analysis of vancomycin in patients with hematological malignancies. Antimicrob Agents Chemother 2005 Dec;49(12):4934-4941 [FREE Full text] [doi: 10.1128/AAC.49.12.4934-4941.2005] [Medline: 16304155]

13. Roberts JA, Taccone FS, Udy AA, Vincent J, Jacobs F, Lipman J. Vancomycin dosing in critically ill patients: robust methods for improved continuous-infusion regimens. Antimicrob Agents Chemother 2011 Jun;55(6):2704-2709. [doi: 10.1128/aac.01708-10]

14. Masich AM, Kalaria SN, Gonzales JP, Heil EL, Tata AL, Claeys KC, et al. Vancomycin pharmacokinetics in obese patients with sepsis or septic shock. Pharmacotherapy 2020 Mar;40(3):211-220. [doi: 10.1002/phar.2367] [Medline: 31957057]

15. Klick Inc and Lakeridge Health. URL: https://www.vancocalc.com [accessed 2022-01-30]

16. Tasa T, Metsvaht T, Kalamees R, Vilo J, Lutsar I. DosOpt: a tool for personalized bayesian dose adjustment of vancomycin in neonates. Ther Drug Monit 2017 Dec;39(6):604-613. [doi: 10.1097/FTD.0000000000000456] [Medline: 29084032]

17. Vancomycin Calculator: Pharmacokinetic calculator with Bayesian modeling. ClinCalc. URL: https://clincalc.com/Vancomycin [accessed 2021-03-19]

18. BestDose. Laboratory of Applied Pharmacokinetics and Bioinformatics. URL: http://www.lapk.org/bestdose.php [accessed 2022-03-09]

19. What is jQuery? jQuery. URL: https://jquery.com [accessed 2021-10-07]

20. Build fast, responsive sites with Bootstrap. Bootstrap. URL: https://getbootstrap.com [accessed 2021-10-07]

21. Chart.js. URL: https://www.chartjs.org [accessed 2021-10-07]

22. Bayes.js. GitHub. URL: https://github.com/rasmusab/bayes.js [accessed 2022-03-16]

23. Statistical methods in readable JavaScript for browsers, servers, and people. Simple Statistics. URL: https://simplestatistics.org/ [accessed 2021-10-07]

24. Sheiner LB, Beal SL. Some suggestions for measuring predictive performance. J Pharmacokinet Biopharm 1981 Aug;9(4):503-512. [doi: 10.1007/BF01060893] [Medline: 7310648]

25. The R Project for Statistical Computing. R Foundation for Statistical Computing. URL: https://www.R-project.org [accessed 2022-03-09]

26. Hiraki Y, Onga T, Mizoguchi A, Tsuji Y. Investigation of the prediction accuracy of vancomycin concentrations determined by patient-specific parameters as estimated by Bayesian analysis. J Clin Pharm Ther 2010 Oct;35(5):527-532. [doi: 10.1111/j.1365-2710.2009.01126.x] [Medline: 20831677]

27. Turner RB, Kojiro K, Shephard EA, Won R, Chang E, Chan D, et al. Review and validation of bayesian dose-optimizing software and equations for calculation of the vancomycin area under the curve in critically ill patients. Pharmacotherapy 2018 Dec;38(12):1174-1183. [doi: 10.1002/phar.2191] [Medline: 30362592]

28. Nunn MO, Corallo CE, Aubron C, Poole S, Dooley MJ, Cheng AC. Vancomycin dosing: assessment of time to therapeutic concentration and predictive accuracy of pharmacokinetic modeling software. Ann Pharmacother 2011 Jun;45(6):757-763. [doi: 10.1345/aph.1P634] [Medline: 21652786]

29. Goti V, Chaturvedula A, Fossler MJ, Mok S, Jacob JT. Hospitalized patients with and without hemodialysis have markedly different vancomycin pharmacokinetics: a population pharmacokinetic model-based analysis. Ther Drug Monit 2018 Apr;40(2):212-221. [doi: 10.1097/FTD.0000000000000490] [Medline: 29470227]

30. Broeker A, Nardecchia M, Klinker KP, Derendorf H, Day RO, Marriott DJ, et al. Towards precision dosing of vancomycin: a systematic evaluation of pharmacometric models for Bayesian forecasting. Clin Microbiol Infect 2019 Oct;25(10):1286.e1-1286.e7 [FREE Full text] [doi: 10.1016/j.cmi.2019.02.029] [Medline: 30872102]

31. Colin PJ, Eleveld DJ, Hart A, Thomson AH. Do vancomycin pharmacokinetics differ between obese and non-obese patients? comparison of a general-purpose and four obesity-specific pharmacokinetic models. Ther Drug Monit 2021 Feb 01;43(1):126-130. [doi: 10.1097/FTD.0000000000000832] [Medline: 33278242]

Abbreviations

AUC: area under the curve
Cl: clearance
IV: intravenous
MedAE: median absolute error
MedPE: median prediction error
PK: pharmacokinetic
RMSE: root mean squared error
TDM: therapeutic drug monitoring
VD: volume of distribution

XSL•FO
RenderX

Original Paper

Mobile App Intervention on Reducing the Myeloproliferative Neoplasm Symptom Burden: Pilot Feasibility and Acceptability Study

Hninyee Win[1], MD; Samantha Russell[1], MD; Betsy C Wertheim[2], MS; Victoria Maizes[3], MD; Robert Crocker[3], MD; Audrey J Brooks[3], PhD; Ruben Mesa[4], MD; Jennifer Huberty[5], PhD; Holly Geyer[4], MD; Ryan Eckert[4], MS, CSCS; Ashley Larsen[1], MS; Krisstina Gowin[5], MD

[1]Department of Medicine, University of Arizona, Tucson, AZ, United States

[2]Cancer Center, University of Arizona, Tucson, AZ, United States

[3]Andrew Weil Center for Integrative Medicine, University of Arizona, Tucson, AZ, United States

[4]Mays Cancer Center, University of Texas, San Antonio, TX, United States

[5]Department of Hematology, University of Arizona, Tucson, AZ, United States

Corresponding Author:
Krisstina Gowin, MD
Department of Hematology
University of Arizona
1501 N Campbell
Tucson, AZ, 85724
United States
Phone: 1 5206260662
Email: gowink@email.arizona.edu

Abstract

Background: Myeloproliferative neoplasms (MPNs) are a group of myeloid malignancies associated with significant symptom burden. Despite pharmacological advances in therapies, inadequate management of MPN symptoms results in reduced quality of life.

Objective: This study aims to determine the feasibility of a 12-week global wellness mobile app intervention in decreasing MPN symptom burden. The University of Arizona Andrew Weil Center for Integrative Medicine's global wellness mobile app, *My Wellness Coach* (MWC), guides patients to improve their health and well-being through facilitating behavior changes.

Methods: Of the 30 patients enrolled in a 12-week intervention, 16 (53%) were retained through the final assessment. Feasibility was assessed by the ease of recruitment, participant adherence, and mobile app acceptability. App acceptability was measured using the user version of the Mobile Application Rating Scale. MPN symptom burden was measured at baseline and 12 weeks after the intervention.

Results: Recruitment was efficient, with the participant goal reached within a 60-day period, suggestive of a demand for such an intervention. Adherence was less than the target within study design (75%), although similar to mobile device app use in other studies (53%). The app was deemed acceptable based on the mean user version of the Mobile Application Rating Scale 3-star rating by participants. Finally, there were statistically significant improvements in several MPN symptoms, quality of life, and total score on the Myeloproliferative Neoplasm Symptom Assessment Form surveys.

Conclusions: Our 12-week intervention with the MWC app was feasible and was associated with a decrease in MPN symptom burden. Further investigation of the MWC app for use as a self-management strategy to reduce the symptom burden in patients with MPN is warranted.

XSL•FO

RenderX

Introduction

Background

Myeloproliferative neoplasms (MPNs) are a group of blood cancers and include the diagnoses of myelofibrosis, essential thrombocythemia, and polycythemia vera. These are rare malignancies with an incidence of 0.33, 1.14, and 1.18 per 100,000 persons per year, respectively [1-3]. Although the prevalence is relatively low, the chronicity of MPNs with significant disease-associated symptoms leads to many years of drastically reduced quality of life for patients afflicted with this disease. Disease manifestations vary among individuals, but generally, MPNs are characterized by excessive blood cells leading to an increased risk of blood clots or bleeding, enlargement of the spleen, and a significant symptom burden.

MPN symptom burden is heterogeneous and may be debilitating to afflicted patients. Fatigue is a major contributor to poor quality of life, affecting more than 80% of patients with MPN. Other symptoms include pruritus, night sweats, difficulty sleeping, abdominal discomfort, bone pain, fever, weight loss, decreased memory, poor sleep, inactivity, and psychosocial dysfunction [4-6]. Although the biology of symptoms is complex, it is rooted to the underlying physiologic inflammation associated with the disease and potentially contributing lifestyle factors and, thus, may be modifiable with intervention [7].

The discovery of the *JAKV617F* mutation in 2005 led to the subsequent development of JAK inhibitor therapy, which has revolutionized the treatment paradigm for MPNs. Despite improvements in disease outcomes offered by pharmacologic therapy, patients' symptoms often remain quite high; thus, strategies to address symptoms represent a significant unmet need in the patient population of MPN [6].

Integrative oncology, defined as the use of complementary and integrative therapies in conjunction with conventional oncology care, may offer a unique symptom management tool [8,9]. Previous publications in integrative oncology have largely focused on breast, colon, and prostate cancers, and there is a scarcity of research to support integrative care in patients with MPN [6,10]. Nevertheless, early data using integrative approaches for the treatment of MPNs are promising, including aerobic activity, yoga, meditation, and strength training, to reduce the symptom burden and improve disease-related inflammation [6,11].

As thrombosis is the primary cause of mortality in MPNs, treatment is largely aimed at reducing thrombotic complications. In addition to pharmacologic treatments, lifestyle modifications through integrative methods may be an important tool in decreasing this risk as well. Inflammation is associated with increased risk of thrombotic events through various mechanisms, including elevated white blood cell and platelet counts, together with activated clotting factors and endothelial cells [12]. Regular exercise reduces inflammatory processes and promotes fibrinolytic activity [13,14]. Studies in patients with underlying cardiovascular disease that have established lifestyle change, including increased physical activity and the subsequent lowering of serum cholesterol, can decrease thrombosis risk [15]. It is possible that this same mechanism can decrease thrombotic risk in patients with MPN.

With the evolution of smartphone technology, mobile apps have become useful tools to help manage chronic diseases and may represent an important strategy for wellness-based interventions [16-24]. In oncology, mobile apps have largely focused on patient education and awareness, including such topics as cancer screening, adverse reactions to treatment, disease processes and treatment options, medication compliance, and cancer pain management [16,17,20,21,24,25]. A few mobile apps have emerged to address the symptom burden in endometrial, breast, sarcoma, and lung cancers with lifestyle changes, mindfulness, and coping mechanisms [18,22,23]. Although smartphone-based meditation and web-based yoga have proven some benefit in patients with MPN, with early data specifically suggesting impact on fatigue, depression, and blood inflammation, there has not been a mobile app to specifically address the MPN symptom burden through promoting global behavioral changes [19,26].

Objectives

The University of Arizona Andrew Weil Center for Integrative Medicine has recently developed and successfully piloted a global wellness mobile app, *My Wellness Coach* (MWC), to guide patients to better health and wellness through facilitating behavior changes [27]. The aim of this study is to determine the feasibility of a global wellness mobile app 12-week intervention for decreasing MPN symptom burden. This is meaningful because this app could provide a highly disseminatable and much needed self-management strategy for the MPN symptom burden.

Methods

Ethics Approval

This study was approved by the University of Arizona Human Subjects Protection Program (IRB Number 2004567060).

Recruitment and Enrollment

Patients with MPN were recruited nationally through organizational partners, such as the Myeloproliferative Neoplasm Quality of Life Group. A flyer was posted on the partners' website, and an email describing the study was sent out through approved listservs, with a link to a web-based eligibility consent and survey over a 60-day period. Through this web-based survey, developed in Qualtrics [28], participants consented and answered questions regarding the eligibility requirements of the study. Participants were screened for eligibility using the inclusion and exclusion criteria given in Textbox 1.

Once a web-based questionnaire was returned, it was evaluated for completeness and eligibility based on the responses. A participant was then emailed the consent form and the baseline questionnaire that were required for enrollment in the study. These participants were given approximately 3 days to respond. If they did not respond within the allocated time, then they were deemed as not interested in participating in the study, and the baseline questionnaire was sent to the next eligible participant.

Eligible participants were contacted until our enrollment goal of 30 individuals was met. An optional phone interview was conducted for qualitative analysis at 12 weeks.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Individuals, man or woman, aged ≥18 years

- Individuals able to provide consent

- Individuals who can speak English and is English literate

- Individuals diagnosed of essential thrombocythemia, polycythemia vera, or myelofibrosis

- Individuals identified by a treating physician

- Individuals who have a personal mobile device with an Android (version 4.1 or higher) or iOS (version 12.2 or higher) operating system

- Individuals who displayed satisfactory ability to operate a mobile device and showcase email communication

- Individuals who have reliable internet and cellular access

Exclusion criteria

- Individuals currently participating in a clinical trial

- Individuals who received blood transfusion within the last 6 months

- Individuals with an Eastern Cooperative Oncology Group [29] status of ≥3

Mobile App Intervention

After consent was obtained from eligible participants, they were provided with instructions to download and use MWC. In the app, the participants were prompted to participate in four steps: (1) score, (2) explore, (3) set a goal, and (4) take action. Self-motivation was emphasized by encouraging reflection on why they want to facilitate change. Clinical research staff suggested setting at least two wellness goals. Participants could set goals within seven domains: nutrition, movement, sleep, resilience, environment, relationships, and spirituality. The app provided tips for health improvements and links to curated resources. When setting goals, participants were coached to make them *SMART*—specific, measurable, attainable, relevant, and time-bound—with the value of this technique described in other publications [30-32]. It provided 24- to 72-hour interval reminders before and after each action step and a goal deadline to encourage action throughout the intervention.

Outcome Measures

Feasibility was assessed by ease of recruitment, adherence, and mobile app rating. Successful recruitment was defined as enrolling 30 participants within 60 days. Adherence was defined as a minimum of 75% retention throughout the intervention. The user version of the Mobile Application Rating Scale (uMARS) evaluated acceptability through rankings in four objective categories (engagement, functionality, aesthetics, and information) and was graded on a Likert scale. This included ease of use, intent for continued use, perceived benefit, and overall star rating.

The Myeloproliferative Neoplasm Symptom Assessment Form (MPN-SAF) evaluated common MPN symptoms [33]. The MPN-SAF comprises 18 symptoms plus a total score. Higher scores indicate worse symptom severity. Each symptom was rated according to the level of difficulty a participant had with each item during the past week, scaled from 0 (absent) to 10 (worst imaginable).

The Behavioral Risk Factor Surveillance System (BRFSS) was used to assess physical activity [34]. The 8-item Patient-Reported Outcomes Information System (PROMIS) Sleep Disturbance questionnaire (Short Form 8a) measured perceptions of sleep quality, depth, and restoration within the past 7 days [35]. Mediterranean diet adherence was assessed using a 14-item questionnaire (which has previously shown an inverse association between the Mediterranean diet and obesity in the PREDIMED [Prevención con Dieta Mediterránea] trial [36]). However, the original 14-point scale was revised because 40% of participants were unable to quantify olive oil consumption and responded "don't know/not sure" or left responses blank. At 12-week follow-up, MPN-SAF, BRFSS, PROMIS Sleep, and Mediterranean diet adherence questionnaires were completed in addition to uMARS. Semistructured phone interviews with the research coordinator to provide additional feedback were optional.

Statistical Analysis

To assess recruitment, acceptability, length of time to accrue, adherence, ease of app use, intent for continued use, and perceived benefit measures were summarized using descriptive statistics. Patient demographics and medical history were summarized using descriptive statistics. Descriptive statistics included proportions for categorical variables and median (IQR) or mean (SD) for continuous variables. Mean changes from baseline to 12 weeks in sleep disturbance, Mediterranean diet adherence, and MPN-SAF symptom scores were tested using linear mixed-effects models, clustered on participants. Such mixed-effects models were used instead of 2-tailed paired t tests to allow all participants (n=30) to contribute to the model, instead of only the subset (n=16) with the end-of-study data. The α level was set to .05, meaning that mean values at 12

weeks were considered significantly different from the mean values at baseline if *P*<.05. All statistical analyses were conducted using Stata (version 16.1; StataCorp). As this is a feasibility study, there were no adjustments made for multiple comparisons because of the small sample size.

Results

Recruitment and Enrollment

Of the 94 interested patients with MPN screened for eligibility, 72 (76%) were eligible to enroll (Figure 1). Emails were sent to eligible patients until 30 were successfully enrolled, which was accomplished within a 3-week period.

Figure 1. Outline of participant enrollment. ECOG: Eastern Cooperative Oncology Group; MPN: myeloproliferative neoplasm.
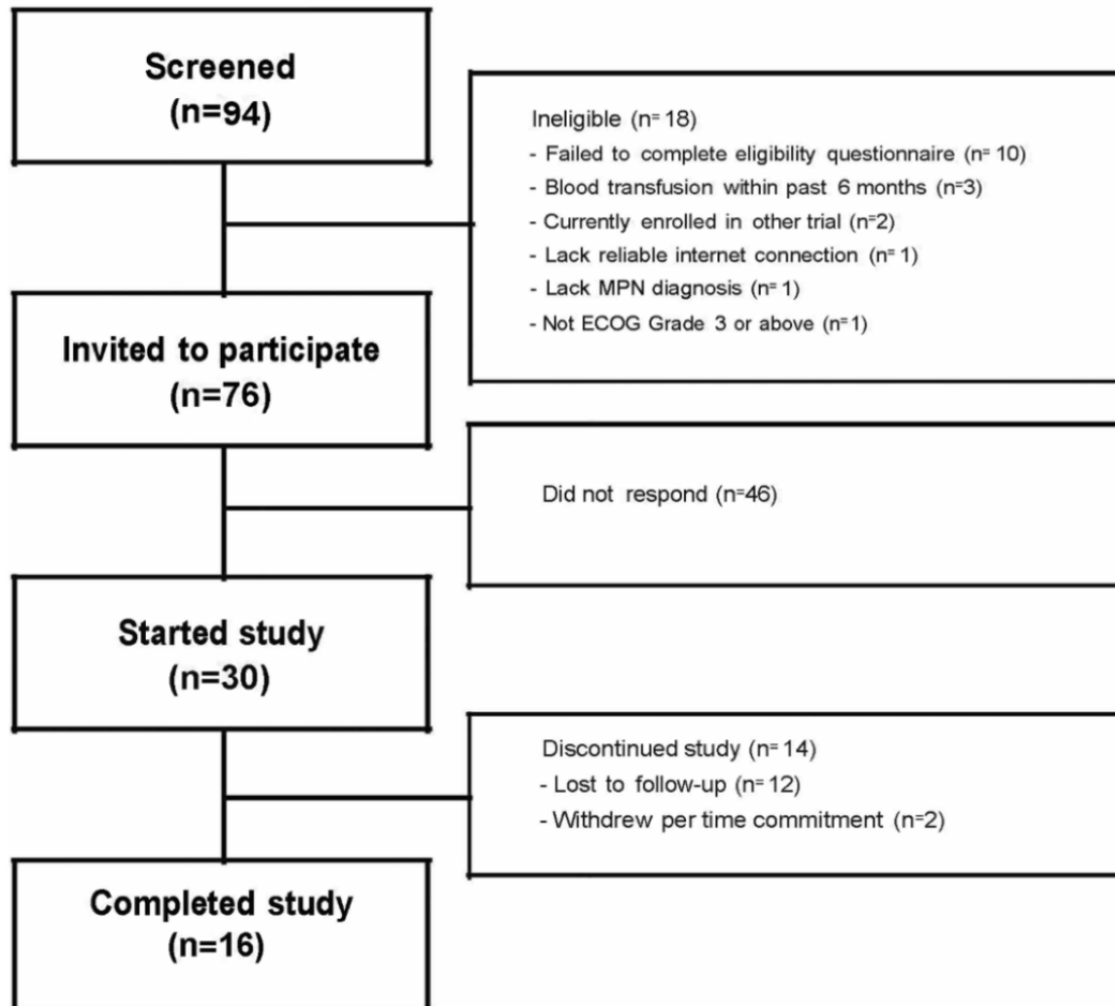
Patient Demographics

The median age of participants was 62.6 (IQR 54.1-70.2; mean 60.5, SD 12.6) years, and participants were predominantly women (26/30, 87%) and White (28/30, 93%). MPN diagnoses included essential thrombocythemia (11/30, 37%), polycythemia vera (11/30, 37%), and myelofibrosis (8/30, 27%). Most participants (23/30, 76%) were diagnosed with MPN >5 years ago. Of the 30 participants, 11 (37%) reported a known history of splenomegaly, 12 (40%) documented a previous diagnosis of anemia, and 6 (20%) reported a history of blood clots after being diagnosed with MPN. Most participants were being treated with hydroxyurea (20/30, 67%), and some (8/30, 27%) were being treated with ruxolitinib or another JAK inhibitor. Of the 12 participants who completed questions pertaining to the Dynamic International Prognostic Scoring system for myelofibrosis, 1 (8%) was in the high-risk group and 6 (50%) were in the intermediate-1–risk group. There were no statistically significant differences in any of the aforementioned characteristics between participants who were retained in the study (16/30, 53%) and those who were not (14/30, 47%; Table 1).

Table 1. Baseline characteristics of participants.

Characteristic	All participants (n=30)	Retained participants (n=16)
Age (years), median (IQR)	62 (54.1-70.2)	62 (54.9-67.5)
BMI (kg/m^2), median (IQR)	24.5 (21.9-29.8)	23 (20.8-34.6)
Gender, n (%)		
Man	3 (10)	2 (13)
Woman	26 (87)	13 (81)
Prefer not to answer	1 (3)	1 (6)
Race, n (%)		
Asian population	1 (3)	1 (6)
White population	28 (93)	15 (94)
Prefer not to identify	1 (3)	0 (0)
Ethnicity, n (%)		
Non-Hispanic or Latino(a) population	25 (96)	14 (100)
Prefer not to identify	1 (4)	0 (0)
Language, n (%)		
English	30 (100)	16 (100)
MPN[a] diagnosis, n (%)		
Essential thrombocytosis	11 (37)	7 (44)
Polycythemia vera	11 (37)	5 (31)
Myelofibrosis	8 (27)	4 (25)
Year of MPN diagnosis, n (%)		
2015 to present	7 (24)	4 (27)
2010-2014	13 (45)	6 (40)
Before 2010	9 (31)	5 (33)
History of enlarged spleen, n (%)	11 (37)	6 (38)
History of anemia, n (%)	12 (40)	4 (25)

[a]MPN: myeloproliferative neoplasm.

MWC Mobile App Use and Rating

Of the 30 enrolled participants, 7 (23%) failed to make an MWC account, and 10 (33%) additional participants did not set wellness goals within the app. Of the remaining 13 participants, 10 (33%) set at least one or two goals (a total of 5 goals allowed). The uMARS questionnaire was successfully completed by 43% (13/30) of the participants. Most categories had a mean score at ≥3 (Table 2). The highest rated category was aesthetics (mean score 3.4, SD 0.8), and the lowest rated category was behavior change (mean score 2.9, SD 1.2). Most participants (7/13, 54%) gave the MWC app an overall 3-star rating acceptable on the 5-point scale [37].

A total of 6 participants provided additional feedback in an optional phone interview. Participants praised that the app addressed "every aspect of healthy living" and found it resourceful. In addition, 3 participants stated that the app encouraged them to work toward achieving their goals. Then, 1 participant specified that it helped her recognize "that there are definite areas that I need to work on and it helped to get a good start working in the right direction."

Most participants (5/6, 83%) felt that additional guidance in setting up goals would be helpful. Other feedback included the frequency of the reminders seemed "too much" and the tone of the reminders could potentially be adjusted.

Table 2. The user version of the Mobile Application Rating Scale (n=13).

Rating	Values
Category, mean (SD; range)	
Engagement	3.0 (1.1; 1.0-5.0)
Functionality	3.2 (1.3; 1.0-5.0)
Aesthetics	3.4 (0.8; 2.0-5.0)
Information	3.2 (1.4; 1.0-5.0)
Overall quality	3.1 (1.1; 1.0-5.0)
Awareness	3.2 (1.3; 1.0-5.0)
Knowledge	3.2 (1.4; 1.0-5.0)
Attitudes	3.2 (1.3; 1.0-5.0)
Intention to change	3.2 (1.4; 1.0-5.0)
Help seeking	3.0 (1.5; 1.0-5.0)
Behavior change	2.9 (1.2; 1.0-5.0)
Overall star rating (single question), n (%)	
1 star	2 (15)
2 stars	2 (15)
3 stars	7 (54)
4 stars	1 (8)
5 stars	1 (8)

Adherence and Acceptability

All 30 participants were successfully enrolled within the allotted 60-day period. Only 53% (16/30) of participants were retained throughout the 12-week study. Attrition was due to either voluntary withdrawal from the study (2/30, 7%) or loss to follow-up for unknown reasons (12/30, 40%). On the basis of the uMARS score as outlined in the *MWC Mobile App Use and Rating* section, MWC was rated a mean score at ≥3 in many categories, with an overall 3-star rating by most of the participants. The baseline characteristics of participants who completed the study and those who did not were compared, and no statistically significant differences were found.

Patient-Reported Outcomes

The MPN-SAF responses highlighted that not all symptoms were changed by the 12-week intervention, but several (8/18, 44%) significantly improved. Improvements were observed in inactivity, impaired concentration, dizziness, numbness, sexual dysfunction, night sweats, bone pain, and quality of life (Table 3). Furthermore, the mean total score was significantly better at 12 weeks than at baseline (mean 54.8, SD 31.6, vs mean 68.4, SD 30.5; *P*=.001).

The PROMIS Sleep Disturbance (Short Form 8a) responses suggested that mean sleep disturbance scores did not significantly change between baseline (mean 24.2, SD 7.2) and at 12 weeks (mean 22.2, SD 7.4; *P*=.49). Few participants reported moderate or severe sleep disturbance at baseline (5/30, 17%) or 12 weeks (3/16, 19%; Table 4). Furthermore, there was a marginally significant (*P*=.06) difference in Mediterranean diet adherence between baseline (mean 5.9, SD 2.3) and at 12 weeks (mean 6.2, SD 2.7).

Of the 16 participants who completed the BRFSS survey, 16 (100%) responded *yes* to *any physical activities or exercises* at baseline compared with 13 (81%) at 12 weeks. The most frequent types of physical activity or exercise included walking or hiking (8/16, 50% participants at baseline; 7/16, 44% at 12 weeks), yoga, Pilates, or barre (3/16, 19% at baseline; 2/16, 13% at 12 weeks), and gardening (2/16, 13% at baseline; 0/16, 0% at 12 weeks). Participants generally reported a high frequency of engagement in their top physical activity, with most participants reporting 5 to 6 times per week (6/16, 38% at baseline; 6/16, 38% at 12 weeks) or ≥7 times per week (4/16, 25% at baseline; 4/16, 25% at 12 weeks).

Table 3. Assessment of symptom severity based on MPN-SAF using a mixed-effects model.

Symptom	Baseline (n=30), mean (SD)	12 weeks (n=16), mean (SD)	β coefficient	P value
Early satiety	4.5 (2.9)	3.8 (3.1)	−0.59	.21
Abdominal pain	2.6 (2.5)	2.8 (2.9)	.10	.80
Abdominal discomfort	3.5 (2.8)	3.6 (3.1)	.01	.99
Inactivity	5.2 (2.6)	3.9 (2.8)	−1.18	.02
Headaches	3.6 (3.0)	2.9 (1.9)	−0.50	.33
Impaired concentration	5.2 (3.0)	3.6 (2.6)	−1.13	.03
Dizziness	4.0 (2.8)	3.1 (2.4)	−1.07	.02
Numbness	3.8 (2.6)	3.0 (2.0)	−0.96	.02
Insomnia	5.9 (3.4)	4.8 (3.1)	−0.53	.20
Depression	3.5 (2.6)	3.2 (2.5)	−0.04	.93
Sexual dysfunction	5.2 (3.8)	3.9 (3.5)	−1.62	.01
Cough	1.9 (1.6)	1.9 (1.9)	−0.23	.50
Night sweats	3.7 (2.5)	2.7 (1.9)	−1.01	.001
Itching	3.7 (3.2)	2.8 (2.7)	−0.05	.84
Bone pain	3.3 (2.8)	2.4 (2.2)	−0.71	.02
Fever	1.3 (1.0)	1.1 (0.2)	.05	.37
Weight loss	2.2 (2.5)	1.4 (0.9)	−0.70	.10
Quality of life	5.3 (2.6)	3.8 (2.2)	−1.27	.03
MPN-SAF[a] total score	68.4 (30.5)	54.8 (31.6)	−11.19	.001

[a]MPN-SAF: Myeloproliferative Neoplasm Symptom Assessment Form.

Table 4. Sleep disturbance based on the Patient-Reported Outcomes Information System.

Level	Baseline (n=29), n (%)	12 weeks (n=16), n (%)
None to slight	14 (48)	10 (63)
Mild	10 (34)	3 (19)
Moderate	4 (14)	3 (19)
Severe	1 (3)	0 (0)

Discussion

Principal Findings

There is a high demand in the patient population of MPN for innovative interventions to help alleviate their symptom burden [4,38]. The aim of this study is to determine the feasibility of a 12-week global wellness mobile-based intervention to reduce symptom burden in patients with MPN. This is the first known study to use a global wellness mobile app to address the symptom burden of patients with MPN through advocating self-management strategies focused on facilitating behavioral change. Our results suggest that this study is feasible in terms of the successful enrollment of 30 MPN participants within 30 days and mobile app acceptability by patients. Previous publications, with many recent additions in the medical literature over the last 2 years, have suggested that mobile health apps provide a feasible option to improve quality of life in various patients with cancer, especially through self-care support and behavior modification [22,23,39-41]. However, no mobile health apps have yet to be tested in the patient population of MPN. Most smartphone apps suggest benefit in symptom management; however, none, to our knowledge, have incorporated symptom management and global lifestyle change within 1 app. In small pilot studies, a smartphone-based meditation (CALM) and web-based yoga have previously been shown to be feasible in the population with MPN, with suggested impact on fatigue, sleep, pain, anxiety, and depression [19,26]. Although these pilot studies suggest benefit in symptom management with an isolated behavior intervention (ie, meditation or yoga), the app tested in our study focused on the feasibility of a global lifestyle intervention. A pilot study of MWC showed similar successful recruitment of patients and simple delivery of a nonpharmacologic intervention [27]. The ease and timeliness of this process suggest that this app may be an effective method for providing a self-management strategy for symptom burden in patients with MPN that is highly disseminable to wide groups of individuals.

With most uMARS mean scores at ≥3 points (of a total possible 5), the mobile app was deemed acceptable to this small group of participants [42]. The lowest rated category was behavior change, with a lower mean score of 2.9, close to the mean score in previous studies. Comparatively, the highest rated category was aesthetics, with a mean score of 3.4. The app could potentially be improved by addressing feedback from participants who stated that the mobile app would benefit from additional information describing the process for goal setting, as well as some technical glitches. Furthermore, our study did not achieve the 75% defined target for adherence, with only 53% (16/30) of the participants retained through the 12-week final assessment. However, this finding is quite common; other studies using mobile device health apps note that most users stop use soon after initial use [43] and that approximately half of the 934 survey responders (45.7%) had stopped using a mobile health app [41]. Another study revealed that only 54% of individuals who download a mobile device health app persisted in using it >1 month [44]. Of the 47% who did not complete the final assessment, most participants were lost to follow-up for unknown causes. Retention in future studies could potentially be improved by having more communication with study coordinators throughout the 12-week duration of the intervention check in and assist with any technical difficulties or questions [45].

MWC aims to provide guidance and facilitate wellness goals. By comparing baseline MPN-SAF assessments with 12-week evaluations, improvement was noted in the areas of inactivity, impaired concentration, dizziness, numbness, sexual dysfunction, night sweats, bone pain, and quality of life. Furthermore, the total score improved after the intervention. Thus, the 12-week intervention with MWC suggests a positive impact on our small study group, with improvement in the overall symptom burden. An app feature that may have contributed to the improvement of the symptom burden is the reminder to act on goals. Previous studies have shown that apps that remind participants have better use rate than apps that did not [44,46]. Given our small group of participants, larger studies will be needed to validate the true impact of MWC on MPN symptoms.

The mobile app's impact on diet did not show significant statistical improvement in Mediterranean diet adherence after the 12-week intervention. Limitations in scoring for Mediterranean diet adherence was noted because of participants leaving certain questions blank or responding as "don't know/not sure." In addition, regarding physical activity, the BRFSS survey results were limited in assessing the quantitative change using statistical methods. Therefore, future studies may need to determine a different method for evaluating physical activity that allows for scoring of physical activity. However, participants were noted to engage in their physical activity of choice at high frequency, with 10 participants engaging 5 to 6 times per week or more. However, the impact on sleep was not statistically significant. Although lifestyle changes were not significantly improved in this pilot study, future analysis with a larger patient sample and potentially alternate dietary and physical activity measures may improve the study design. Poor cardiometabolic health is a primary contributor to mortality in patients with MPN, and effective lifestyle interventions and continued study of lifestyle interventions are needed.

Overall, the study exhibited ease of recruitment and acceptability, and although adherence was lower than our goal, it was commensurate with other mobile health app interventions. Finally, the 12-week mobile app pilot study improved the MPN symptom burden and warrants continued study with a larger sample size.

Limitations

The most notable limitation to our study was the small study size. Not achieving the defined level of adherence is a concern; however, the 75% target was likely set too high when compared with the mobile health app use found in the literature and, therefore, may point to a limitation of study design [41,43]. In addition, only 53% (16/30) of the participants completed the study, and 47% (14/30) voluntarily withdrew or were lost to follow-up. On the basis of the completion results in this study, future studies could potentially increase the recruitment goal to ensure a larger sample completing the intervention.

Additional limitations of the study include variability in the proficiency of modern technology by the participants and inability to collect data regarding weekly use of the app. Some participants admitted to discontinuing the mobile app because of confusion and difficulty in understanding how to properly use the app. This limitation could be surmounted by future study design, including frequent coordinator and group engagement for app support. Participants also expressed concerns regarding the frequency of reminders, with a participant stating that it was too frequent and another participant stating that she was not reminded enough. These limitations highlight the importance of a user-friendly mobile app platform with the ability for customization. Perhaps letting the users set their own frequency for reminders can improve adherence. In addition, the study was conducted during the COVID-19 pandemic, when gym closures, self-isolation, and mental health issues could confound the participants' sleep, diet, and physical activity. A participant specifically stated that she "would have used the app for better success if [she] hadn't been self-isolating during this pandemic."

Access barriers, such as patients with MPN who do not possess a mobile device, reliable internet connection, or limited literacy, should also be acknowledged. Although no adaptations for the older patient population were incorporated in this feasibility study, users have the ability within the app to increase the font size on their mobile device for visual limitations. In addition, users can dictate narrative responses (voice-to-text functionality), which can reduce the typing required. App adaptations for the older population will be considered in future design and study of the mobile app. Finally, this study did not investigate which mechanisms were most effective at reducing specific symptoms, as this would require a much larger sample size, and this should be explored in future studies.

Conclusions

Our small feasibility pilot study provided preliminary evidence that the MWC global wellness app intervention is feasible within the population with MPN and may reduce the MPN symptom burden. There was insufficient evidence to show improvements

in physical activity vigor and frequency, healthy diet consumption, and better sleep quality. Future trials evaluating the impact of the MWC mobile app in the population with MPN could be strengthened by a larger sample size, measures to improve adherence and retention (such as coordinator support), and closer consideration of diet and physical activity measures. Overall, this study provides a promising first step toward a self-management strategy to lessen the substantial symptom burden of patients with MPN.

Conflicts of Interest

RM is a consultant for Novartis, Sierra Onc, La Jolla Pharma, and Constellation.

References

1. Mehta J, Wang H, Iqbal SU, Mesa R. Epidemiology of myeloproliferative neoplasms in the United States. Leuk Lymphoma 2014 Mar;55(3):595-600. [doi: 10.3109/10428194.2013.813500] [Medline: 23768070]
2. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 2016 May 19;127(20):2391-2405 [FREE Full text] [doi: 10.1182/blood-2016-03-643544] [Medline: 27069254]
3. Shallis RM, Wang R, Davidoff A, Ma X, Podoltsev NA, Zeidan AM. Epidemiology of the classical myeloproliferative neoplasms: the four corners of an expansive and complex map. Blood Rev 2020 Jul;42:100706. [doi: 10.1016/j.blre.2020.100706] [Medline: 32517877]
4. Mesa RA, Niblack J, Wadleigh M, Verstovsek S, Camoriano J, Barnes S, et al. The burden of fatigue and quality of life in myeloproliferative disorders (MPDs): an international Internet-based survey of 1179 MPD patients. Cancer 2007 Jan 01;109(1):68-76 [FREE Full text] [doi: 10.1002/cncr.22365] [Medline: 17123268]
5. Lin A, Kase JA, Moore MM, Son I, Tran N, Clotilde LM, et al. Multilaboratory validation of a luminex microbead-based suspension array for the identification of the 11 most clinically relevant Shiga toxin-producing Escherichia coli O serogroups. J Food Prot 2013 May;76(5):867-870. [doi: 10.4315/0362-028X.JFP-12-468] [Medline: 23643130]
6. Surapaneni P, Scherber RM. Integrative approaches to managing myeloproliferative neoplasms: the role of nutrition, exercise, and psychological interventions. Curr Hematol Malig Rep 2019 Jun;14(3):164-170. [doi: 10.1007/s11899-019-00516-w] [Medline: 31093888]
7. Geyer HL, Dueck AC, Scherber RM, Mesa RA. Impact of inflammation on myeloproliferative neoplasm symptom development. Mediators Inflamm 2015;2015:284706 [FREE Full text] [doi: 10.1155/2015/284706] [Medline: 26538823]
8. Deng GE, Rausch SM, Jones LW, Gulati A, Kumar NB, Greenlee H, et al. Complementary therapies and integrative medicine in lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest 2013 May;143(5 Suppl):420-436. [doi: 10.1378/chest.12-2364] [Medline: 23649450]
9. Lyman GH, Greenlee H, Bohlke K, Bao T, DeMichele AM, Deng GE, et al. Integrative therapies during and after breast cancer treatment: ASCO endorsement of the SIO clinical practice guideline. J Clin Oncol 2018 Sep 01;36(25):2647-2655. [doi: 10.1200/JCO.2018.79.2721] [Medline: 29889605]
10. Lin W, Zhong M, Zhou Q, Zhang Y, Wang H, Zhao Z, et al. Efficacy of complementary and integrative medicine on health-related quality of life in cancer patients: a systematic review and meta-analysis. Cancer Manag Res 2019;11:6663-6680 [FREE Full text] [doi: 10.2147/CMAR.S195935] [Medline: 31413628]
11. Gowin K, Langlais BT, Kosiorek HE, Dueck A, Millstine D, Huberty J, et al. The SIMM study: survey of integrative medicine in myeloproliferative neoplasms. Cancer Med 2020 Dec;9(24):9445-9453 [FREE Full text] [doi: 10.1002/cam4.3566] [Medline: 33140580]
12. Moliterno AR, Ginzburg YZ, Hoffman R. Clinical insights into the origins of thrombosis in myeloproliferative neoplasms. Blood 2021 Mar 04;137(9):1145-1153. [doi: 10.1182/blood.2020008043] [Medline: 33237986]
13. Chen Y, Apostolakis S, Lip GYH. Exercise-induced changes in inflammatory processes: implications for thrombogenesis in cardiovascular disease. Ann Med 2014 Nov;46(7):439-455. [doi: 10.3109/07853890.2014.927713] [Medline: 25012964]
14. Martin K. Risk factors for and management of MPN-associated bleeding and thrombosis. Curr Hematol Malig Rep 2017 Oct;12(5):389-396. [doi: 10.1007/s11899-017-0400-3] [Medline: 28948496]
15. Shah PK, Lecis D. Inflammation in atherosclerotic cardiovascular disease. F1000Res 2019;8 [FREE Full text] [doi: 10.12688/f1000research.18901.1] [Medline: 31448091]
16. Bae WK, Kwon J, Lee HW, Lee S, Song E, Shim H, et al. Feasibility and accessibility of electronic patient-reported outcome measures using a smartphone during routine chemotherapy: a pilot study. Support Care Cancer 2018 Nov;26(11):3721-3728. [doi: 10.1007/s00520-018-4232-z] [Medline: 29732483]

XSL•FO

RenderX

17. Cruz FO, Vilela RA, Ferreira EB, Melo NS, Reis PE. Evidence on the use of mobile apps during the treatment of breast cancer: systematic review. JMIR Mhealth Uhealth 2019 Aug 27;7(8):e13245 [FREE Full text] [doi: 10.2196/13245] [Medline: 31456578]

18. Donovan E, Martin SR, Seidman LC, Zeltzer LK, Cousineau TM, Payne LA, et al. A mobile-based mindfulness and social support program for adolescents and young adults with sarcoma: development and pilot testing. JMIR Mhealth Uhealth 2019 Mar 18;3(3):e10921 [FREE Full text] [doi: 10.2196/10921] [Medline: 30882352]

19. Huberty J, Eckert R, Dueck A, Kosiorek H, Larkey L, Gowin K, et al. Online yoga in myeloproliferative neoplasm patients: results of a randomized pilot trial to inform future research. BMC Complement Altern Med 2019 Jun 07;19(1):121 [FREE Full text] [doi: 10.1186/s12906-019-2530-8] [Medline: 31174535]

20. Jibb LA, Stevens BJ, Nathan PC, Seto E, Cafazzo JA, Johnston DL, et al. Implementation and preliminary effectiveness of a real-time pain management smartphone app for adolescents with cancer: a multicenter pilot clinical study. Pediatr Blood Cancer 2017 Oct;64(10). [doi: 10.1002/pbc.26554] [Medline: 28423223]

21. Kim HJ, Kim SM, Shin H, Jang J, Kim YI, Han DH. A mobile game for patients with breast cancer for chemotherapy self-management and quality-of-life improvement: randomized controlled trial. J Med Internet Res 2018 Oct 29;20(10):e273 [FREE Full text] [doi: 10.2196/jmir.9559] [Medline: 30578205]

22. McCarroll ML, Armbruster S, Pohle-Krauza RJ, Lyzen AM, Min S, Nash DW, et al. Feasibility of a lifestyle intervention for overweight/obese endometrial and breast cancer survivors using an interactive mobile application. Gynecol Oncol 2015 Jun;137(3):508-515. [doi: 10.1016/j.ygyno.2014.12.025] [Medline: 25681782]

23. Park S, Kim JY, Lee JC, Kim HR, Song S, Kwon H, et al. Mobile phone app-based pulmonary rehabilitation for chemotherapy-treated patients with advanced lung cancer: pilot study. JMIR Mhealth Uhealth 2019 Feb 04;7(2):e11094 [FREE Full text] [doi: 10.2196/11094] [Medline: 30714943]

24. Wyatt KD, Jenkins SM, Plevak MF, Venegas Pont MR, Pruthi S. A personalized, web-based breast cancer decision making application: a pre-post survey. BMC Med Inform Decis Mak 2019 Oct 21;19(1):196 [FREE Full text] [doi: 10.1186/s12911-019-0924-7] [Medline: 31638964]

25. Silveira CE, Carcano C, Mauad EC, Faleiros H, Longatto-Filho A. Cell phone usefulness to improve the skin cancer screening: preliminary results and critical analysis of mobile app development. Rural Remote Health 2019 Jan;19(1):4895 [FREE Full text] [doi: 10.22605/RRH4895] [Medline: 30673294]

26. Huberty J, Eckert R, Larkey L, Kurka J, De Jesús SA, Yoo W, et al. Smartphone-based meditation for myeloproliferative neoplasm patients: feasibility study to inform future trials. JMIR Form Res 2019 Apr 29;3(2):e12662 [FREE Full text] [doi: 10.2196/12662] [Medline: 31033443]

27. Yuan N, Brooks AJ, Burke MK, Crocker R, Stoner GM, Cook P. My wellness coach: evaluation of a mobile app designed to promote integrative health among underserved populations. Andrew Weil Center for Integrative Medicine (Under Review). URL: https://lifepathsresearch.org/wp-content/uploads/2021/09/MWC-Pilot_APHA-2019-poster.pdf [accessed 2022-03-02]

28. Qualtrics XM. 2021. URL: https://www.qualtrics.com/au/?rid=ip&prevsite=en&newsite=au&geo=IN&geomatch=au [accessed 2022-03-02]

29. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 1982 Dec;5(6):649-655. [Medline: 7165009]

30. Bowman J, Mogensen L, Marsland E, Lannin N. The development, content validity and inter-rater reliability of the SMART-Goal Evaluation Method: a standardised method for evaluating clinical goals. Aust Occup Ther J 2015 Dec;62(6):420-427. [doi: 10.1111/1440-1630.12218] [Medline: 26286379]

31. Bailey RR. Goal setting and action planning for health behavior change. Am J Lifestyle Med 2019 Sep 13;13(6):615-618 [FREE Full text] [doi: 10.1177/1559827617729634] [Medline: 31662729]

32. Strecher VJ, Seijts GH, Kok GJ, Latham GP, Glasgow R, DeVellis B, et al. Goal setting as a strategy for health behavior change. Health Educ Q 1995 May;22(2):190-200. [doi: 10.1177/109019819502200207] [Medline: 7622387]

33. Emanuel RM, Dueck AC, Geyer HL, Kiladjian J, Slot S, Zweegman S, et al. Myeloproliferative neoplasm (MPN) symptom assessment form total symptom score: prospective international assessment of an abbreviated symptom burden scoring system among patients with MPNs. J Clin Oncol 2012 Nov 20;30(33):4098-4103 [FREE Full text] [doi: 10.1200/JCO.2012.42.3863] [Medline: 23071245]

34. Desmond R, Jackson BE, Hunter G. Utilization of 2013 BRFSS Physical Activity Data for state cancer control plan objectives: Alabama data. South Med J 2015 May;108(5):290-297 [FREE Full text] [doi: 10.14423/SMJ.0000000000000282] [Medline: 25972217]

35. Hanish AE, Lin-Dyken DC, Han JC. PROMIS sleep disturbance and sleep-related impairment in adolescents: examining psychometrics using self-report and actigraphy. Nurs Res 2017;66(3):246-251 [FREE Full text] [doi: 10.1097/NNR.0000000000000217] [Medline: 28448375]

36. Martínez-González MA, García-Arellano A, Toledo E, Salas-Salvadó J, Buil-Cosiales P, Corella D, et al. A 14-item Mediterranean diet assessment tool and obesity indexes among high-risk subjects: the PREDIMED trial. PLoS One 2012;7(8):e43134 [FREE Full text] [doi: 10.1371/journal.pone.0043134] [Medline: 22905215]

XSL•FO
RenderX

37. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]

38. Rossau HK, Kjerholt M, Brochmann N, Tang LH, Dieperink KB. Daily living and rehabilitation needs in patients and caregivers affected by myeloproliferative neoplasms (MPN): a qualitative study. J Clin Nurs 2021 Jul 06. [doi: 10.1111/jocn.15944] [Medline: 34231273]

39. Gustavell T, Sundberg K, Segersvärd R, Wengström Y, Langius-Eklöf A. Decreased symptom burden following surgery due to support from an interactive app for symptom management for patients with pancreatic and periampullary cancer. Acta Oncol 2019 Sep;58(9):1307-1314. [doi: 10.1080/0284186X.2019.1633473] [Medline: 31284797]

40. Stubbins R, He T, Yu X, Puppala M, Ezeana CF, Chen S, et al. A behavior-modification, clinical-grade mobile application to improve breast cancer survivors' accountability and health outcomes. JCO Clin Cancer Inform 2018 Dec;2:1-11 [FREE Full text] [doi: 10.1200/CCI.18.00054] [Medline: 30652617]

41. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]

42. Berry DL, Blonquist TM, Patel RA, Halpenny B, McReynolds J. Exposure to a patient-centered, web-based intervention for managing cancer symptom and quality of life issues: impact on symptom distress. J Med Internet Res 2015;17(6):e136 [FREE Full text] [doi: 10.2196/jmir.4190] [Medline: 26041682]

43. Lee K, Kwon H, Lee B, Lee G, Lee JH, Park YR, et al. Effect of self-monitoring on long-term patient engagement with mobile health applications. PLoS One 2018;13(7):e0201166 [FREE Full text] [doi: 10.1371/journal.pone.0201166] [Medline: 30048546]

44. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]

45. Gupta L, Khandelwal D, Lal PR, Gupta Y, Kalra S, Dutta D. Factors determining the success of therapeutic lifestyle interventions in diabetes - role of partner and family support. Eur Endocrinol 2019 Apr;15(1):18-24 [FREE Full text] [doi: 10.17925/EE.2019.15.1.18] [Medline: 31244906]

46. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inform Syst 2009;24. [doi: 10.17705/1CAIS.02428]

Abbreviations

BRFSS: Behavioral Risk Factor Surveillance System
MPN: myeloproliferative neoplasms
MPN-SAF: Myeloproliferative Neoplasm Symptom Assessment Form
MWC: My Wellness Coach
PREDIMED: Prevención con Dieta Mediterránea
PROMIS: Patient-Reported Outcomes Information System
uMARS: user version of the Mobile Application Rating Scale

Original Paper

Patient Experience and Satisfaction in Online Reviews of Obstetric Care: Observational Study

Emily K Seltzer[1,2], MPH; Sharath Chandra Guntuku[1,3], PhD; Amy L Lanza[1], MPH; Christopher Tufts[1], MS; Sindhu K Srinivas[4], MD, MSCE; Elissa V Klinger[1,2], MS; David A Asch[2,5], MD, MBA; Nick Fausti[1], BA; Lyle H Ungar[3], PhD; Raina M Merchant[1,2], MD, MSHP

[1]Penn Medicine Center for Digital Health, University of Pennsylvania, Philadelphia, PA, United States

[2]Penn Medicine Center for Health Care Innovation, University of Pennsylvania, Philadelphia, PA, United States

[3]Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, United States

[4]Department of Obstetrics and Gynecology, University of Pennsylvania, Philadelphia, PA, United States

[5]Center for Health Equity Research and Promotion, Corporal Michael J. Crescenz VA Medical Center, Philadelphia, PA, United States

Corresponding Author:
Sharath Chandra Guntuku, PhD
Department of Computer and Information Science
University of Pennsylvania
3330 Walnut St
Philadelphia, PA, 19103
United States
Phone: 1 215 615 3211
Email: sharathg@seas.upenn.edu

Abstract

Background: The quality of care in labor and delivery is traditionally measured through the Hospital Consumer Assessment of Healthcare Providers and Systems but less is known about the experiences of care reported by patients and caregivers on online sites that are more easily accessed by the public.

Objective: The aim of this study was to generate insight into the labor and delivery experience using hospital reviews on Yelp.

Methods: We identified all Yelp reviews of US hospitals posted online from May 2005 to March 2017. We used a machine learning tool, latent Dirichlet allocation, to identify 100 topics or themes within these reviews and used Pearson r to identify statistically significant correlations between topics and high (5-star) and low (1-star) ratings.

Results: A total of 1569 hospitals listed in the American Hospital Association directory had at least one Yelp posting, contributing a total of 41,095 Yelp reviews. Among those hospitals, 919 (59%) had at least one Yelp rating for labor and delivery services (median of 9 reviews), contributing a total of 6523 labor and delivery reviews. Reviews concentrated among 5-star (n=2643, 41%) and 1-star reviews (n=1934, 30%). Themes strongly associated with favorable ratings included the following: top-notch care ($r=0.45$, $P<.001$), describing staff as comforting ($r=0.52$, $P<.001$), the delivery experience ($r=0.46$, $P<.001$), modern and clean facilities ($r=0.44$, $P<.001$), and hospital food ($r=0.38$, $P<.001$). Themes strongly correlated with 1-star labor and delivery reviews included complaints to management ($r=0.30$, $P<.001$), a lack of agency among patients ($r=0.47$, $P<.001$), and issues with discharging from the hospital ($r=0.32$, $P<.001$).

Conclusions: Online review content about labor and delivery can provide meaningful information about patient satisfaction and experiences. Narratives from these reviews that are not otherwise captured in traditional surveys can direct efforts to improve the experience of obstetrical care.

KEYWORDS

XSL•FO
RenderX

Introduction

Many hospitals in the United States use the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) and Press Ganey surveys to evaluate patient experiences [1]. Survey results are standardized and publicly reported to facilitate comparisons of patient experience. However, they are costly and often have low response rates [2,3]. Prespecified domains may miss the concerns of many patients, and aggregated public reporting can obscure differences across specialties [4,5].

Yelp is a website where users share information about their experiences at local businesses by giving a star rating from 1-5 and leaving a narrative review. Yelp is the most used free website in the United States for hospital ratings [6]. In one study, 65% of gynecologists reported being likely to use online ratings to improve patient care, more so than physicians from other specialties [7]. Prior work demonstrates that reviews from online rating services like Yelp are correlated with traditional methods for understanding the patient experience, and the platform's unstructured design provides information not captured in conventional patient experience surveys [2,8-12]. The scale and utilization of these platforms is significant and may provide a

nuanced way to better listen to patients [13]. In this study, we aim to evaluate how the content of labor and delivery Yelp reviews relates to star rating to provide insight into the labor and delivery experience in the United States.

Methods

Obtaining Hospital Reviews

We identified hospitals in the United States that have Yelp reviews using the Yelp Search application programming interface. We included only hospitals listed in the American Hospital Association directory with at least one review. Hospital reviews were then searched for keywords specific to labor and delivery, identified by referencing the Unified Medical Language System database and gathering input from an obstetrician (SKS). The search terms included variations of the same word—for example, "deliver" and "delivery" were both used but counted as one search term. Reviews containing at least one of the specified keywords were characterized as "labor and delivery reviews" and all others as "non–labor and delivery" (Table 1). We used only reviews that received a 5-star or 1-star review for the final analyses, considering the bimodal distribution (Table 2).

Table 1. Characteristics of hospitals with labor and delivery Yelp reviews.

Characteristic	All hospitals on Yelp (N=1569)		All labor and delivery hospitals on Yelp (N=919)	
	Hospitals, n (%)	Average Yelp rating (on a scale from 1-5)	Hospitals, n (%)	Average Yelp rating (on a scale from 1-5)
Bed size				
0-49	223 (14)	3.23	46 (5)	3.44
50-199	577 (37)	2.86	329 (36)	3.31
200-399	487 (31)	2.84	334 (36)	3.24
≥400	275 (18)	2.88	210 (23)	3.33
Region				
Northeast	305 (20)	2.84	191 (21)	3.16
South	617 (40)	2.90	330 (36)	3.23
Midwest	184 (12)	3.12	71 (8)	3.45
West	456 (29)	2.88	327 (36)	3.35
Teaching hospital				
Yes	192 (12)	2.92	157 (17)	3.27
No	1370 (88)	2.85	762 (83)	3.29

Table 2. Distribution of ratings of hospital reviews describing labor and delivery versus all hospital reviews.

Rating (stars)	Labor and delivery Yelp reviews (N=6523), n (%)	Hospital Yelp reviews (N=41,095), n (%)
1	1957 (30)	16,849 (41)
2	522 (8)	3288 (8)
3	457 (7)	2466 (6)
4	913 (14)	5342 (13)
5	2674 (41)	13,150 (32)

Deriving Language Features

After removing stop words, common words of low information content (eg, "the," "as," "a"), we used the MALLET implementation [14] of the machine learning program latent Dirichlet allocation (LDA) to generate 100 topics based on prior work [13]. This machine learning technique automates the identification of co-occurring words whose combination suggests themes or topics [15]. For example, the frequent co-occurrence of "hours," "waiting," "sitting," and "lobby" would define a topic which, on inspection, suggests the theme of long wait times. LDA was used to build a topic model using the corpus of review text; afterward, each review was represented as a weighted mixture of the 100 topics generated from the reviews.

Identifying Differentially Expressed Language Features

Our analysis was aimed at identifying differentially expressed topics in reviews with a 1-star (low) rating versus a 5-star (high) rating considering the bimodal distribution of ratings and based on prior work [11,12]. All statistical analyses were performed in R (version 3.4.1; R Foundation for Statistical Computing). We took a data-driven approach to allow for a more transparent view of the words and phrases that differentiate posts with a high rating (5-star) from those with a low rating (1-star). We isolated the patterns in language topics to obtain correlations in both groups using ordinary least squares (OLS) regression. Treating each review as an observation, OLS regression was performed on standardized LDA derived variables for each review, with the reviews that received 5 stars labeled as 1 and those that received 1 star labeled as 0, and the LDA topic weights of the written review text as the independent variables. Since the variables were standardized, the OLS regression coefficients can be interpreted as Pearson correlations. Topics with a positive coefficient are therefore associated with 5-star reviews, and topics with large negative coefficients are associated with 1-star reviews. We used Bonferroni correction and $P<.001$ for indicating meaningful correlations and the effect size was measured using Pearson r. Most highly correlated topics were labeled independently by two coauthors by examining the top 7 terms in each topic. Adjudication of discrepancies occurred via consensus with a third coauthor reviewer.

Ethical Considerations

The University of Pennsylvania Institutional Review Board deemed the study exempt.

Results

Hospital Reviews

We identified 41,095 reviews from 1569 hospitals listed in the American Hospital Association directory with at least one Yelp rating posted from May 2005 to March 2017. Among those hospitals, 919 (59%) had at least one Yelp rating for labor and delivery (median of 9), contributing a total of 6523 labor and delivery reviews about labor and delivery services. The distribution of ratings is shown in Table 2.

Differentially Expressed Language Features

Themes correlated with favorable ratings included the following: top-notch care ($r=0.45$), expressing gratitude toward staff ($r=0.41$), describing staff as comforting ($r=0.52$), staff having good bedside manner ($r=0.42$), professional and friendly staff ($r=0.43$), the delivery experience ($r=0.46$), modern and clean facilities ($r=0.44$), and hospital food ($r=0.38$; Table 3).

Themes correlated with 1-star labor and delivery reviews included the experience of calling the hospital ($r=0.33$), interactions with reception ($r=0.31$), complaints to management ($r=0.30$), telling others to avoid the hospital ($r=0.32$), a lack of agency among patients ($r=0.47$), and issues with discharging from the hospital ($r=0.32$; Table 4).

Table 3. Yelp differential language analysis topics associated with positive (5-star) labor and delivery reviews.

Yelp domain (determined by Yelp) and Yelp topic (topic terms)	Correlation, Pearson r	Example quote
Positive experience		
Top-notch care (excellent, received, care, top, notch, wonderful, attentive, amazing)	0.448	Thank you to the ENTIRE [hospital name] Pediatric unit. They have taken EXCELLENT care of our baby. Your attention and dedication was top notch. We greatly appreciate it....
Compassionate caring staff		
Grateful (amazing, team, wonderful, grateful, life, love, god, team, saved, bless, remember)	0.405	We are very blessed to have such an amazing team of nurses. The support and caring was priceless. We most likely will have baby 2 in here. If you looking for a happy medium between home-birth or hospital-birth, this place is the answer.
Comforting (made, comfortable, feel, helpful, questions, friendly, caring, pleasant, attentive)	0.521	Both of my daughters gave birth at [the hospital] and raved about the care they got. The nursing staff is compassionate and very skilled!
Bedside manner (great, bedside, dr, manner, awesome, kind, fantastic, sweet, tech, compassionate)	0.416	Delivered my first here. Clean, friendly, knowledgeable staff. Very attentive in Labor and Delivery and until we went home. Made the week long stay as comfortable as possible. Nurses had really good bedside manner. I would recommend this hospital to others.
Professional/friendly staff (friendly, staff, professional, helpful, efficient, service, recommend, highly, courteous)	0.426	Excellent labor/delivery and postpartum experience. Every nurse we encountered was kind and caring. Highly recommended.
Clinical service		
Delivery experience (delivery, labor, baby, amazing, birth, wonderful, maternity, nicu[a], ward, helpful)	0.462	Maternity ward was awesome when I delivered my baby in October everyone was exceptional.
Facilities and amenities		
Modern and clean (facility, rooms, clean, friendly, helpful, nice, super, equipment, modern, beautiful)	0.444	This is hospital is one of the cleanest around! The maternity staff is excellent and so are the facilities!
Hospital food (food, pretty, order, stay, cafeteria, menu, private, dinner, free)	0.380	I delivered my daughter here and absolutely love everything about this hospital from the warm staff, comfortable rooms and amazing kitchen menu for patient meals.

[a]NICU: neonatal intensive care unit.

XSL•FO
RenderX

Table 4. Yelp differential language analysis topics associated with negative (1-star) labor and delivery reviews.

Yelp domain (determined by Yelp) and Yelp topic (topic terms)	Correlation, Pearson r	Example quote
Negative interactions		
Lack of agency (told, asked, didn't, questions, wasn't, couldn't, rude, talk, telling, upset)	0.465	This review is for the prenatal clinic. I called this morning because I had been advised to do so by my Dr. The lady who answered the phone was extremely rude and unprofessional. Her exact words " and your calling us for????" With the rudest tone I have ever heard from a healthcare professional. I was beyond shocked and will never go to their clinic.
Being discharged (information, discharge, medications, papers, husband, refused, stated, physician, attending, signed)	0.321	Beware!!! Don't entrust this incompetent facility with your life!! Intake form - misspelling of name and incorrect recording of birth date so records could not be found until 4phone calls later…
Communication with hospital		
Calls (phone, call, number, person, message, answer, hold, office, transferred)	0.326	Horrible customer service. Called the operator to inquire about setting up a prenatal appointment to find an OBGYN and she told me to google them to find one I like! Ha what a joke.
Reception (desk, front, asked, check, walked, minutes, arrived, paperwork)	0.305	Maternity receptionist is still rude as hell. Wife sent over by office for emergency monitoring.... She told us to wait while she finished putting stickers on a folder.
Complaints to management (complaint, manager, response, letter, lack, contact, advocate, concerns, case, report)	0.296	I had a baby at [hospital name] in 2008. Horrible, scary, TRAUMATIC experience. Incompetence, unprofessionalism, and bad medicine…In spite of numerous conversations with various individuals at the hospital, and a lengthy grievance filed against the hospital with my insurance, no one at the hospital ever said "I'm sorry."
Wait times		
Long wait (hours, waited, hour, finally, worst, sitting, triage, ridiculous)	0.409	Anyone giving this place 5 stars works here. Longest wait ever over 2 hours with a ectopic pregnancy. This place is horrible
Pain meds		
Medication (pain, meds, gave, prescription, morphine, prescribed, migraine, severe)	0.325	If you want to be constantly asked if you're a drug addict while you cry in pain this is your place. Even at 43 under a Sutter West doctors care, had a baby there 3 months earlier have 9 years of [hospital name] medical records, and that still didn't help the staff treat not me like an addict until my x-Rays came back then it's like I'm so sorry here's some pills have a nice day call your doctor. :(
Would not recommend		
Unprofessional staff (worst, rude, terrible, awful, avoid, incompetent, horrible, place, unprofessional)	0.315	This Hospital is one of the worst hospitals in the [county name] county. Their staff is very rude. They almost killed my mother during child birth due to uneducated staff and negligence. DO NOT GO HERE! Not safe.
Specialized medical care		
Tests (test, results, sample, urine, lab, ordered, UTI[a], negative)	0.342	I came in for my 8week old baby he had a temp. Of 101F I came in at 11pm and now it's 4:15 am we are still waiting on some results I know it does not take that long for something to come back about my baby knowing the baby haves a high temp.
IV[b] (arm, hand, needle, IV, put, started, painful, stop, screaming, catheter, veins)	0.339	Never ever would I bring my baby to this hospital walked in my room and found a needle on the floor
Kidney stones (kidney, pain, abdominal, severe, stomach, stone, meds, excruciating)	0.315	I went in for serious stomach pains while I was pregnant, the doctor said I had gall stones that needed to be removed via surgery, scheduled surgery for after my baby was born just to find out later that there was no gall stones, never did find out what sent me to the emergency room in debilitating pain while I was pregnant!

[a]UTI: urinary tract infection.

[b]IV: intravenous.

XSL•FO

RenderX

Discussion

Principal Findings

This study found identifiable themes associated with high and low ratings, offering insights into what patients seeking labor and delivery services care about most. Online reviews about hospitals include comments about the experience of labor and delivery care. Although online reviews are not validated and may attract or amplify the most negative comments [13,16], they reflect raw reports from patients unconstrained by pre-established topics.

Positive reviews on the labor and delivery experience overwhelmingly cited compassionate and attentive hospital staff. Nurses were frequently cited as the most important component of the experience. For 5-star reviewers who criticized their experience in any way, caring and helpful nurses and staff almost always made up for the negative aspects of their stay. In addition, 5-star reviews in our study largely referenced positive feelings about hospital staff and the importance of hospital amenities (often citing spa showers, advanced technology, and appealing decor). Prior work reported that, in the patient-provider relationship in an obstetrics and gynecology setting, patients reported greater satisfaction with their health care experience when they had a positive relationship with their care team, which parallels our finding that patients are more satisfied when providers are caring and attentive [17]. Compassion of staff is not a topic measured in HCAHPS surveys. Additionally, HCAHPS and Press Ganey do not include free-text questions; rather, questions are multiple choice.

Negative reviews of labor and delivery included topics typically inverse to the topics discussed in positive reviews. Raters cited negative interactions and lack of communication with hospital staff, long wait times, and low-quality obstetrics care in 1-star labor and delivery reviews. In a prior review of patient satisfaction in obstetrics care, researchers interviewed patients and compiled a total of 51 items related to patient satisfaction [18]. The list included multiple characteristics related to provider communication style, including compassion/sensitivity, communication, accessibility, support, and positive affirmation of birthing process. Access to and communication with hospital staff contribute to a more positive patient experience in the context of labor and delivery care. Understanding the common themes of positive and negative experiences may help clinical and operational staff create initiatives and protocols that lead to better patient encounters.

Limitations

This study has several limitations. The American Hospital Association data set represents broader obstetrical programs (eg, the Hospital of the University of Pennsylvania) but may miss subsidiary programs (eg, Penn Ob/Gyn & Midwifery Care). The bimodal distribution of reviews may amplify the voices of those with strongly positive and strongly negative experiences, muting the more nuanced and mixed experiences. Clinical terms and procedures may be talked about in slang and ways that are harder to identify using automated techniques. However, using machine learning techniques allows for the analysis of hundreds of thousands of reviews as opposed to what is possible with human coders. In addition, Yelp reviews are not validated and may vary in quality and quantity. To counter this, we eliminated reviews "not recommended" by Yelp (a measure indicating a review is likely to be fake). "Not recommended" reviews are determined automatically by Yelp's proprietary algorithm that considers a number of factors to try and remove fake reviews (eg, one person posting many reviews from the same computer). In the future, including other online review platforms may provide richer insights. The practical application of this data is largely valuable as a supplemental insight into the patient's psychological experience of their labor and delivery care. Understanding the themes that correlate to high and low reviews may provide a place to start when developing standardized surveys for measuring care.

Conclusions

Transparency of hospital performance data is vital to enhancing patient trust and improving health care delivery. Online rating websites may help foster trust and goodwill between hospitals and their consumers, allow consumers to make more informed decisions, and encourage quality improvements [16,19]. Increasing the validity and scientific rigor of these narrative feedback platforms may increase the value of these patient narratives for further improving obstetrics care in the United States [20,21].

Conflicts of Interest

DAA is a partner and part owner of VAL Health, and is a US government employee. All other authors declare no conflicts of interest.

References

1. Zusman EE. HCAHPS replaces Press Ganey survey as quality measure for patient hospital experience. Neurosurgery 2012 Aug;71(2):N21-N24. [doi: 10.1227/01.neu.0000417536.07871.ed] [Medline: 22811203]
2. Ranard BL, Werner RM, Antanavicius T, Schwartz HA, Smith RJ, Meisel ZF, et al. Yelp Reviews Of Hospital Care Can Supplement And Inform Traditional Surveys Of The Patient Experience Of Care. Health Aff (Millwood) 2016 Apr;35(4):697-705 [FREE Full text] [doi: 10.1377/hlthaff.2015.1030] [Medline: 27044971]

XSL•FO
RenderX

3.   Jordan H, White A, Joseph C. Costs and benefits of HCAHPS: final report. Abt Assoc Inc. URL: https://tinyurl.com/2p8jjz5p [accessed 2021-06-22]

4.   The HCAHPS survey - frequently asked questions. 2015. URL: https://tinyurl.com/542f3894 [accessed 2019-04-04]

5.   O'Malley AJ, Zaslavsky A, Hays R, Hepner KA, Keller S, Cleary PD. Exploratory factor analyses of the CAHPS Hospital Pilot Survey responses across and within medical, surgical, and obstetric services. Health Serv Res 2005 Dec;40(6 Pt 2):2078-2095 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00471.x] [Medline: 16316439]

6.   Bardach NS, Asteria-Peñaloza R, Boscardin WJ, Dudley RA. The relationship between commercial website ratings and traditional hospital performance measures in the USA. BMJ Qual Saf 2013 Mar 23;22(3):194-202 [FREE Full text] [doi: 10.1136/bmjqs-2012-001360] [Medline: 23178860]

7.   Emmert M, Meszmer N, Sander U. Do Health Care Providers Use Online Patient Ratings to Improve the Quality of Care? Results From an Online-Based Cross-Sectional Study. J Med Internet Res 2016 Sep 19;18(9):e254 [FREE Full text] [doi: 10.2196/jmir.5889] [Medline: 27644135]

8.   Bardach NS, Lyndon A, Asteria-Peñaloza R, Goldman LE, Lin GA, Dudley RA. From the closest observers of patient care: a thematic analysis of online narrative reviews of hospitals. BMJ Qual Saf 2016 Dec;25(11):889-897 [FREE Full text] [doi: 10.1136/bmjqs-2015-004515] [Medline: 26677215]

9.   Kilaru AS, Meisel ZF, Paciotti B, Ha YP, Smith RJ, Ranard BL, et al. What do patients say about emergency departments in online reviews? A qualitative study. BMJ Qual Saf 2016 Jan 24;25(1):14-24. [doi: 10.1136/bmjqs-2015-004035] [Medline: 26208538]

10. Woods KD. Search Your Heart: What Press Ganey Can Never Measure. Obstet Gynecol 2015;126:7. [doi: 10.1097/aog.0000000000000949]

11. Agarwal AK, Pelullo AP, Merchant RM. "Told": the Word Most Correlated to Negative Online Hospital Reviews. J Gen Intern Med 2019 Jul 14;34(7):1079-1080 [FREE Full text] [doi: 10.1007/s11606-019-04870-6] [Medline: 30767117]

12. Agarwal AK, Mahoney K, Lanza AL, Klinger EV, Asch DA, Fausti N, et al. Online Ratings of the Patient Experience: Emergency Departments Versus Urgent Care Centers. Ann Emerg Med 2019 Jun;73(6):631-638. [doi: 10.1016/j.annemergmed.2018.09.029] [Medline: 30392737]

13. Merchant RM, Volpp KG, Asch DA. Learning by Listening-Improving Health Care in the Era of Yelp. JAMA 2016 Dec 20;316(23):2483-2484. [doi: 10.1001/jama.2016.16754] [Medline: 27997663]

14. McCallum AK. Mallet: A machine learning for language toolkit. MALLET. URL: https://mimno.github.io/Mallet/ [accessed 2021-06-22]

15. Blei D, Ng A, Jordan M. Latent Dirichlet Allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text] [doi: 10.5555/944919.944937]

16. Lee V. Transparency and Trust — Online Patient Reviews of Physicians. N Engl J Med 2017 Jan 19;376(3):197-199. [doi: 10.1056/nejmp1610136]

17. Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. Lancet 2001 Mar 10;357(9258):757-762. [doi: 10.1016/s0140-6736(00)04169-6] [Medline: 11253970]

18. Howell EA, Concato J. Obstetric patient satisfaction: asking patients what they like. Am J Obstet Gynecol 2004 Jan;190(1):175-182. [doi: 10.1016/j.ajog.2003.06.006] [Medline: 14749656]

19. Lagu T, Metayer K, Moran M, Ortiz L, Priya A, Goff SL, et al. Website Characteristics and Physician Reviews on Commercial Physician-Rating Websites. JAMA 2017 Feb 21;317(7):766-768 [FREE Full text] [doi: 10.1001/jama.2016.18553] [Medline: 28241346]

20. Cognetta-Rieke C, Guney S. Analytical Insights from Patient Narratives: The Next Step for Better Patient Experience. J Patient Exp 2014 May;1(1):20-22 [FREE Full text] [doi: 10.1177/237437431400100105] [Medline: 28725797]

21. Schlesinger M, Grob R, Shaller D, Martino SC, Parker AM, Finucane ML, et al. Taking Patients' Narratives about Clinicians from Anecdote to Science. N Engl J Med 2015 Aug 13;373(7):675-679. [doi: 10.1056/NEJMsb1502361] [Medline: 26267629]

Abbreviations

HCAHPS: Hospital Consumer Assessment of Healthcare Providers and Systems
LDA: latent Dirichlet allocation
OLS: ordinary least squares

XSL•FO
RenderX

XSL•FO
RenderX

Original Paper

A Lifestyle Intervention to Delay Early Chronic Kidney Disease in African Americans With Diabetic Kidney Disease: Pre-Post Pilot Study

Mukoso N Ozieh[1*], MD, MSCR; Leonard E Egede[1*], MD, MS

Department of Medicine, Division of Nephrology, Medical College of Wisconsin, Milwaukee, WI, United States
[*]all authors contributed equally

Corresponding Author:
Mukoso N Ozieh, MD, MSCR
Department of Medicine
Division of Nephrology
Medical College of Wisconsin
8701 Watertown Plank Road
Milwaukee, WI, 53226
United States
Phone: 1 4149558839
Email: mozieh@mcw.edu

Abstract

Background: Behavioral factors, such as lifestyle, have been shown to explain approximately 24% of the excess risk of chronic kidney disease (CKD) among African Americans. However, there are limited intervention studies culturally tailored to African Americans with type 2 diabetes mellitus and CKD.

Objective: The main objective of this study was to examine the feasibility and preliminary efficacy of a culturally tailored lifestyle intervention among African Americans with type 2 diabetes mellitus and CKD.

Methods: A pre-post design was used to test the feasibility of a lifestyle intervention in 30 African American adults recruited from the Medical University of South Carolina between January 2017 and February 2017. A research nurse delivered the manualized study intervention weekly for 6 weeks. Clinical outcomes (hemoglobin A_{1c}, blood pressure, and estimated glomerular filtration rate [eGFR]) were measured at baseline and postintervention. Disease knowledge, self-care, and behavior outcomes were also measured using validated structured questionnaires at baseline and postintervention. Descriptive statistics and effect sizes were calculated to determine clinically important changes from baseline.

Results: Significant pre-post mean differences and decreases were observed for hemoglobin A_{1c} (mean 0.75%, 95% CI 0.16-1.34; P=.01), total cholesterol (mean 16.38 mg/dL, 95% CI 5.82-26.94; P=.004), low-density lipoprotein (mean 13.73 mg/dL, 95% CI 3.91-23.54; P=.008), and eGFR (mean 6.73 mL/min/1.73m^2, 95% CI 0.97-12.48; P=.02). Significant pre-post mean differences and increases were observed for CKD self-efficacy (mean −11.15, 95% CI −21.55 to −0.75; P=.03), CKD knowledge (mean −2.62, 95% CI −3.98 to −1.25; P<.001), exercise behavior (mean −1.21, 95% CI −1.96 to −0.46; P=.003), and blood sugar testing (mean −2.15, 95% CI −3.47 to −0.83; P=.003).

Conclusions: This study provides preliminary data for a large-scale appropriately powered randomized controlled trial to examine a culturally tailored lifestyle intervention in African Americans with type 2 diabetes mellitus and CKD in order to improve clinical, knowledge, self-care, and behavior outcomes in this population.

KEYWORDS

XSL·FO
RenderX

Introduction

Chronic kidney disease (CKD), categorized as stage 1 to 5 based on the estimated glomerular filtration rate (eGFR), is a major complication of diabetes and is commonly referred to as diabetic kidney disease (DKD) [1]. DKD is marked by persistent presence of albuminuria (albumin excretion rate ≥30 mg/24 hours; urine albumin to creatinine ratio [UACR] ≥30 mg/g) or a decreased eGFR (<60 mL/min/1.73m^2). Approximately 31% of individuals with diabetes have DKD [2]. DKD is associated with significant morbidity, cost, a 4- to 5-fold risk of end-stage renal disease, and an increased risk of death [3-6]. African Americans are disproportionately affected by diabetes and DKD compared with non-Hispanic Whites [7]. Further, African Americans are 3 to 4 times more likely to have end-stage renal disease compared with non-Hispanic Whites [4]. The reason for the accelerated progression of kidney disease among African Americans is not completely understood. However, behavioral factors, such as lifestyle, have been shown to explain approximately 24% of the excess risk of CKD among African Americans [8].

Lifestyle modification comprising one or more aspects, including physical activity, diet change, smoking cessation, exercise, skills training, counseling, and stress management, is an essential component of diabetes and CKD management [9-11]. Evidence of the impact of lifestyle interventions on DKD outcomes is limited and conflicting [12,13]. The largest prospective study until date, the Look AHEAD (Action for Health in Diabetes) study, randomly assigned overweight or obese patients with type 2 diabetes to an intensive lifestyle intervention to achieve weight loss compared with a diabetes support and education condition [14,15]. While the intervention was not effective in reducing cardiovascular events (primary outcome), a post-hoc analysis of this study showed that the intensive lifestyle intervention reduced the incidence of very high–risk CKD [15].

The majority of lifestyle intervention studies in patients with DKD are limited by their study design, limited generalizability, small sample size, low proportion of African Americans, or lack of appropriate predefined renal endpoints [12,13,15,16]. In addition, there is a lack of interventions culturally tailored to African Americans, even though evidence suggests that African Americans have a limited understanding of CKD and CKD risk factors [17,18], and lag behind in hemoglobin A_{1c} (HbA$_{1c}$) control, blood pressure control, and use of statins and glucose-lowering medications [19]. Hence, the main objective of this pilot study was to examine the feasibility and preliminary efficacy of a culturally tailored DKD-focused lifestyle intervention on (1) clinical outcomes, (2) disease knowledge, and (3) self-care and behavior outcomes among African Americans with type 2 diabetes and CKD. The study hypothesized that individuals who receive the study intervention will have improved clinical outcomes, disease knowledge, and self-care and behavior outcomes after the intervention.

Methods

Ethics Board Review

This study was approved by the Institutional Review Board of the Medical University of South Carolina (Pro#00051414; Institutional Review Board approval date: November 9, 2016).

Study Design

A pre-post design was used to test the feasibility of a lifestyle intervention with baseline (preintervention) and 2-month (postintervention) assessments. The study participants were non-Hispanic Blacks with type 2 diabetes and CKD, with an eGFR >59 mL/min/1.73m^2 and a spot UACR of 30-300 mg/g.

Participants and Setting

Participants were recruited from clinics affiliated with the Medical University of South Carolina between January 2017 and February 2017. Non-Hispanic Black participants were identified and recruited using clinic billing records for ICD-10 (International Classification of Diseases, 10th revision) codes consistent with the diagnosis of type 2 diabetes and cystatin C eGFR >59 mL/min/1.73m^2, and through referral from physicians and clinic staff. Institutional Review Board–approved study flyers were posted in the clinics, and letters of invitation signed by the clinic director were mailed to patients.

Screening for Eligibility and Enrollment

Individuals who were aged 21 years or older, self-identified as African American, had a clinical diagnosis of type 2 diabetes and early CKD (stage 1 and 2), were able to communicate in English, and had a telephone (landline or cell phone) were eligible to participate in the study. Individuals who had cognitive impairment, alcohol or drug abuse, acute decompensation of chronic disease conditions, CKD of stage 3 or higher, malignancy, life-expectancy of less than 6 months, and other known disease conditions causing proteinuria were excluded from the study. Transplant recipients, individuals participating in another diabetes or CKD trial, and those who did not have telephone access were not eligible for the study.

A total of 30 participants who met the inclusion criteria were enrolled by a research assistant. Eligible participants received up to US $150 in compensation for completing all the study assessments (screening, baseline, and 2-month study assessment).

Description of the Intervention

The study intervention was adapted from a culturally tailored study, Technology-Intensified Diabetes Education and Skills Training Intervention (TIDES) [20], and was tailored to focus on DKD. The study intervention was based on the Information-Motivation-Behavioral Skills model and provides information, motivation, and behavioral skills training (using motivational enhancement techniques) [21]. Patients were assigned the FORA 2-in-1 Telehealth System at the beginning of the study and provided glucose test strips to allow testing at least once a day.

All intervention sessions were telephone-delivered weekly by a research nurse for 6 weeks. Intervention sessions lasted 30

minutes, and the research nurse was trained in behavioral skills counseling and the study intervention content. The study educational materials were developed based on the National Kidney Disease Education Program [22] and written in lay language for African Americans with DKD. The description of weekly content is provided in Multimedia Appendix 1. All education sessions were audiotaped, and 20% were randomly selected and reviewed by the principal investigator to ensure the research nurse delivered the intervention appropriately. All study participants received behavioral skills training focused on 3 lifestyle behaviors (physical activity, diet, and medication adherence). Target lifestyle behavior goals were set in collaboration with the patients and were guided by current problem areas and preferences.

Study Measures and Data Collection Schedule

Participant information was collected using validated questionnaires administered at 2 time points: at baseline and 2 months postintervention (see Multimedia Appendix 2). Study data were obtained by a trained research assistant.

Feasibility Measures

Feasibility measures were recruitment, session attendance rate, and dropout proportion.

Outcome Measures

BMI was calculated using weight in kg and height in m^2. Blood pressure readings were obtained using automated blood pressure monitors at baseline and 2 months. The device was programmed to take 3 readings at 2-minute intervals, and the readings were averaged. UACR was measured at baseline and 2-month visits using spot urine. Blood samples were assayed for HbA_{1c}, cholesterol, and eGFR at baseline and 2 months by a trained nurse. The Patient Health Questionnaire-9 (PHQ-9), a brief questionnaire that scores each of the 9 DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, fourth edition) criteria, was used to assess for depression [23]. See Multimedia Appendix 2 for details on outcome and process measures.

Process and Behavioral Measures

CKD Self-efficacy Scale

This involved a 25-item instrument that measures disease-related self-efficacy in the following 4 core areas: (1) autonomy, (2) self-regulation, (3) problem solving, and (4) seeking social support [24]. The Cronbach alpha coefficient for the total scale was .94, and the value for each of the 4 subscales ranged from .84 to .90 [24].

CKD Knowledge Questionnaire

This involved a 28-item Kidney Knowledge survey, which has good internal consistency and high reliability (coefficient of 0.72) [25].

Diabetes Knowledge Questionnaire

This involved the 24-item Diabetes Knowledge Questionnaire, which has a reliability coefficient of 0.78 [26].

Health Literacy

This was measured using the 3-item Chew health literacy scale, which assesses the capacity to obtain, process, and understand basic health-related decisions [27]. The 3 questions have been shown to be effective in detecting inadequate health literacy (areas under the receiver operating characteristic curve of 0.87, 0.80, and 0.76, respectively) [27].

Behavioral Skills

This was assessed with the Summary of Diabetes Self-Care Activities (SDSCA) scale [28]. It is a brief, validated, self-report questionnaire of diabetes self-management that includes items assessing diet, exercise, medication adherence, and self-blood glucose testing. The average interitem correlations within scales were high, test-retest correlations were moderate, and correlations with other measures of diet and exercise generally supported the validity of the subscales.

Statistical Analyses

Important measures of feasibility analysis included recruitment, session attendance rate, and dropout proportion. We used 95% CIs for proportions to estimate (1) the proportion of participants who agreed to participate among those who were initially approached, (2) the proportion who were compliant with the treatment intervention, and (3) the proportion who dropped out. In addition, frequency distributions describing the participants' reasons for noncompliance and discontinuation of study participation will be provided.

For quantitative analysis, univariate descriptive statistics and frequency distributions were calculated for the total sample. Pre-post mean differences were tested using paired t tests. Effect size, a measure of treatment effect, was used to interpret the effects of the intervention. An effect size of 0.2 was considered small, 0.5 was considered moderate, and 0.8 was considered large. In addition to effect sizes, which demonstrated clinical relevance, a statistically significant difference was noted for the primary measurement. All statistical analyses were performed using Stata software (StataCorp).

Results

Study Profile

Between January 9, 2017, and April 28, 2017, 77 patients were screened, and 30 eligible patients were enrolled into the study (Figure 1). All 30 (100%) patients completed the baseline assessment, and 26 (87%) completed assessments at 2 months. Four (13%) participants were lost to follow-up; hence, the analytical sample included 26 participants.

Figure 1. Study flow diagram.
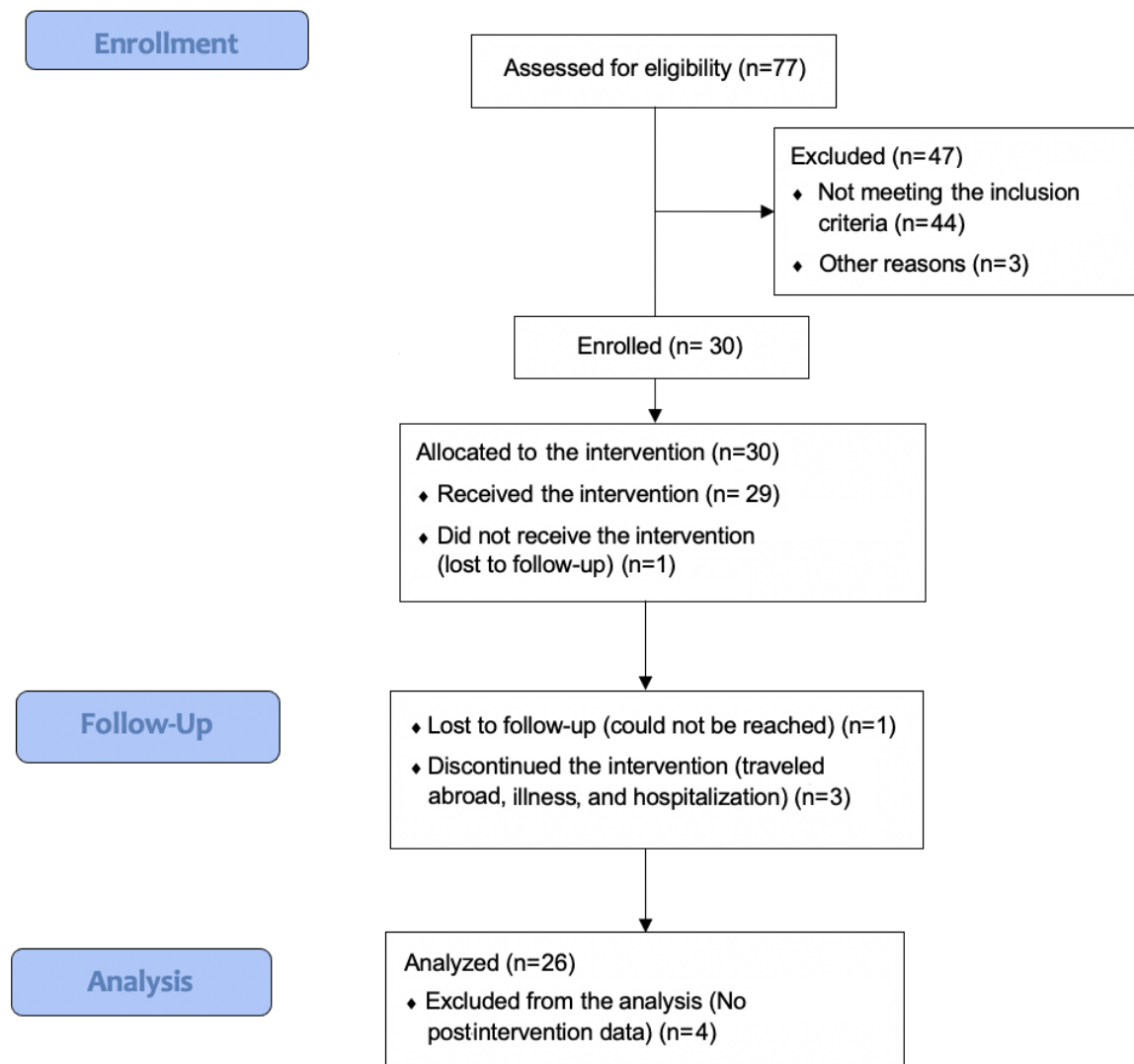
Baseline Demographic Profile

Table 1 shows the baseline characteristics of the study participants. The mean age of the study participants was 57 years, and the mean duration of diabetes was 14 years. The majority were female (21/30, 70%), unmarried (17/30, 57%), unemployed (20/30, 67%), and insured (30/30, 100%). Over half of the participants reported having a "good" health status (16/30, 53%) and not using any special equipment (16/30, 53%).

Table 1. Baseline sample demographic characteristics of African Americans with type 2 diabetes and chronic kidney disease enrolled in the pre-post study.

Variable	Value (N=30)
Age (years), mean (SD)	56.7 (13.5)
Sex, n (%)	
Female	21 (70)
Marital status, n (%)	
Married	13 (43)
Employment status, n (%)	
Employed	10 (33)
Income, n (%)	
<US $25,000	18 (60)
Smoking status, n (%)	
Smoker	9 (30)
Insurance status, n (%)	
Insured	30 (100)
Health status, n (%)	
Fair/poor	14 (47)
Use of special equipment, n (%)	
Yes	14 (47)
Physical activity days per week, n (%)	
0	16 (53)
1+	14 (47)
Years of education, mean (SD)	12.6 (1.7)
Work hours per week, mean (SD)	14.3 (24.0)
Duration of diabetes (years), mean (SD)	14.2 (9.1)

Feasibility Findings

Overall, 33 out of 77 (43%) participants contacted were eligible for the study, and 30 participants were successfully recruited for this study. Among the 30 participants, 21 (70%) completed all 6 sessions, 26 (87%) completed assessments at 2 months, and 1 (3%) dropped out of the study. The reasons for incomplete sessions were death in the family, illness, or hospitalization. One participant dropped out after enrollment because of hospitalization for pneumonia.

Preintervention and Postintervention Differences in Clinical Outcomes

Table 2 presents preintervention (baseline) and postintervention (2 months) differences in clinical outcomes. Significant preintervention and postintervention mean differences and decreases were observed for HbA_{1c} (mean 0.75; P=.01), total cholesterol (mean 16.38; P=.004), low-density lipoprotein (LDL) (mean 13.73; P=.008), and eGFR (mean 6.73; P=.02). We observed nonstatistically significant increases in BMI (mean −0.48; P=.05), systolic blood pressure (mean −1.77; P=.61), diastolic blood pressure (mean −3.42; P=.21), and the UACR (mean −18.63; P=.79). There were also nonstatistically significant decreases in the PHQ-9 score for depression (mean 1.30; P=.17), high-density lipoprotein (mean 1.70; P=.23), and triglycerides (mean 4.03; P=.62).

Table 2. Preintervention and postintervention mean differences in the clinical outcomes of African Americans with type 2 diabetes and chronic kidney disease enrolled in the pre-post study.

Variable	Baseline[a], mean (SD)	2 months[a], mean (SD)	Mean difference (95% CI)	P value
BMI (kg/m^2)	35.9 (8.1)	36.4 (8.4)	−0.48 (−0.96 to 0.01)	.05
Systolic blood pressure (mmHg)	131.2 (17.9)	133.0 (15.2)	−1.77 (−8.84 to 5.30)	.61
Diastolic blood pressure (mmHg)	77.0 (10.2)	80.5 (9.5)	−3.42 (−7.80 to 0.96)	.12
Hemoglobin A_{1c} (%)	9.1 (2.5)	8.3 (1.9)	0.75 (0.16 to 1.34)	.01
Total cholesterol (mg/dL)	199.5 (44.9)	183.1 (34.6)	16.38 (5.82 to 26.94)	.004
Low-density lipoprotein (mg/dL)	122.2 (41.4)	108.4 (31.1)	13.73 (3.91 to 23.54)	.008
High-density lipoprotein (mg/dL)	50.4 (11.0)	48.7 (12.8)	1.70 (−1.14 to 4.52)	.23
Triglycerides (mg/dL)	134.3 (58.3)	130 (61.6)	4.03 (−12.80 to 20.88)	.62
Microalbumin-creatinine ratio	264.8 (375.5)	283.5 (557.6)	−18.63 (−168.83 to 131.58)	.79
Glomerular filtration rate (mL/min/1.73m^2)	93.4 (21.9)	86.7 (21.7)	6.73 (0.97 to 12.48)	.02
Patient Health Questionnaire-9 (PHQ-9) score	5.2 (4.9)	3.9 (4.5)	1.30 (−0.62 to 3.23)	.17

[a]Baseline is preintervention and 2 months is postintervention.

Preintervention and Postintervention Differences in Knowledge, Self-care, and Behavior Outcomes

Table 3 displays participant responses to questions related to knowledge, skills, self-care, and behavior outcomes. Significant preintervention and postintervention mean differences and increases were observed for CKD self-efficacy (mean −11.15; P=.03), CKD knowledge (mean −2.62; P<.001), exercise behavior (mean −1.21, P=.003), and blood sugar testing (mean −2.15; P=.003). We observed nonstatistically significant increases in diet (mean −0.42; P=.49), special diet (mean −0.43; P=.23), and foot care (mean −0.21; P=.60). A nonstatistically significant decrease in diabetes knowledge (mean 0.77; P=.25) was also observed, while no significant change in health literacy (mean 0.04; P=.66) was seen.

Table 3. Pre-post mean differences in knowledge, self-care, and behavior outcomes of African Americans with type 2 diabetes and chronic kidney disease enrolled in the pre-post study.

Variable	Baseline[a], mean (SD)	2 months[a], mean (SD)	Mean difference (95% CI)	P value
Knowledge				
CKD[b] self-efficacy	208.9 (36.0)	220.0 (27.4)	−11.15 (−21.55 to −0.75)	.03
DM[c] knowledge	16.2 (3.8)	15.4 (3.7)	0.77 (−0.59 to 2.14)	.25
CKD knowledge	18.4 (3.8)	21.0 (2.1)	−2.62 (−3.98 to −1.25)	<.001
Health literacy	0.3 (0.5)	0.3 (0.5)	0.04 (−0.14 to −0.21)	.66
Self-care and behavior				
Diet	3.9 (2.1)	4.4 (2.1)	−0.42 (−1.68 to 0.83)	.49
Special diet	4.0 (1.4)	4.4 (1.2)	−0.43 (−1.16 to 0.30)	.23
Exercise behavior	1.8 (1.7)	3.0 (1.6)	−1.21 (−1.96 to −0.46)	.003
Blood sugar test	3.7 (2.6)	5.9 (1.8)	−2.15 (−3.47 to −0.84)	.003
Foot care	4.9 (2.2)	5.1 (2.1)	−0.21 (−1.05 to 0.63)	.60

[a]Baseline is preintervention and 2 months is postintervention.

[b]CKD: chronic kidney disease.

[c]DM: type 2 diabetes.

Discussion

Principal Findings

This study examined the feasibility and preliminary efficacy of a culturally tailored DKD-focused lifestyle intervention in African Americans with type 2 diabetes and CKD. With 100% recruitment, a 70% session attendance rate, and a 3% drop-out rate, the study findings suggest that the design, recruitment, and delivery of a culturally tailored lifestyle intervention for high-risk African Americans with type 2 diabetes and CKD are feasible. This study was also designed to examine preliminary changes in clinical outcomes, disease knowledge, self-care, and

XSL•FO

RenderX

behavior outcomes. We observed statistically significant changes in the clinical outcomes of HbA_{1c}, total cholesterol, LDL, and eGFR following the study intervention. In addition, there were statistically significant increases in CKD self-efficacy, CKD knowledge, exercise. and blood sugar testing.

Comparison With Prior Work

Behavior lifestyle intervention trials have conflicting results on the impact of lifestyle interventions on clinical outcomes [12]. Consistent with our study findings, a systematic review by Van Huffel et al evaluating the impact of exercise and diet on health outcomes in individuals with diabetes and CKD concluded that exercise and diet interventions have beneficial effects on glycemic control, BMI, and quality of life [13]. Similarly, large trials, such as the "Reduction of Endpoints in NIDDM with the Angiotensin II Antagonist Losartan (RENAAL)," "Action to Control Cardiovascular Risk in Diabetes (ACCORD)," "Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT)," and "Action in Diabetes and Vascular Disease: Preterax and Diamicron Modified Release Controlled Evaluation (ADVANCE)," demonstrated that aggressive risk factor control in African Americans using antihypertensive, antihyperglycemic, or lipid-lowering medications is beneficial [29-33]. However, these studies focused on risk factor control using medications and did not emphasize lifestyle modification, which is a core component of diabetes and CKD management [3,9,10].

Contrary to our study findings, a systematic review and meta-analysis of self-management support interventions for people with diabetes and CKD showed that these interventions may improve self-care activities, HbA_{1c}, and systolic blood pressure [34]. While we observed a significant increase in CKD self-efficacy, disease knowledge, exercise behavior, and blood sugar testing, our study did not show a significant difference in blood pressure; however, it was not powered to confirm or refute a hypothesis, which could explain the lack of statistical significance in most clinical outcomes. The impacts of lifestyle interventions on kidney function are also inconsistent, with some studies demonstrating no effect, or a negative or positive effect [12,13]. We observed a significant negative effect (decrease) in the eGFR postintervention in our study population. Glomerular hyperfiltration often mediated by hyperglycemia results in a high eGFR in type 2 diabetes and is a hallmark finding in DKD [35]. Nonpharmacological interventions, such as decreases in body weight, and salt and protein intake, have been shown to ameliorate diabetic hyperfiltration [35]. It is unclear why we observed these findings given the lack of a significant difference in BMI or dietary habits. Future large-scale and more rigorous behavior lifestyle randomized controlled trials in this population should explore measuring changes in salt and protein intake, and examine the impact on outcomes.

Recent evidence demonstrates that glucose-lowering medications, such as glucagon-like peptide 1 receptor agonists and sodium-glucose cotransporter-2 inhibitors, are of particular benefit in the prevention and treatment of CKD in patients with type 2 diabetes [36]. However, despite the strengths of these large clinical trials [37-44] and limited data on the efficacy of lifestyle interventions in African Americans [12,15,34], African Americans remain poorly represented. It is established that African Americans with CKD are poorly represented in clinical trials [45], and low inclusion of African Americans in clinical trials limits the generalizability of study findings. This potentially propagates existing disparities in a high-risk high-cost population. Low participation of African Americans in clinical trials is often attributed to poverty, lack of accessibility, lack of information on clinical trials, and chronic disease–related stigma [45,46]. There is a need to overcome these barriers and increase the participation of African Americans in clinical trials. Ongoing clinical trials are exploring novel community-based screening recruitment methods for African Americans with CKD [46,47]. More intervention studies that focus on high-risk patients incorporating such novel recruitment strategies are needed. In addition, behavioral lifestyle interventions that account for contextual factors facing high-risk African American populations with diabetes and CKD are needed [48].

Strengths, Limitations, and Future Direction

The findings of this study are promising and have important clinical implications. Significant changes observed in clinical outcomes, such as a decrease in HbA_{1c}, and improved CKD knowledge, self-care, and behavior, can prevent or delay the progression of CKD to renal failure, and improve quality of life and survival in this study population. This could potentially reduce the economic burden associated with renal failure and the life-threatening complications of renal failure. Despite these promising findings, some limitations are worth noting. First, the relatively small sample size, limited intervention duration, and lack of a control group might have affected the findings. However, the goal of this feasibility pilot study was to generate information needed for planning and designing a future large-scale study. Second, eGFR was estimated using creatinine and cystatin C equations with race. Recent evidence suggests that the inclusion of race in eGFR estimation overestimates measured eGFR, which potentially exacerbates health disparities and contributes to systemic racism. While it is unlikely that the eGFR equation used for this study influenced the study findings, future studies will use new creatinine and cystatin C equations without race to ensure accuracy. Third, although majority of the study participants completed all intervention sessions, some weekly intervention sessions were delayed. The main reasons for delayed intervention sessions were travel abroad, hospitalization, and death in the family. Future studies will incorporate a run-in period to establish expectations and processes for timely completion of intervention sessions in the event of hospitalization or unanticipated events. In addition, we will account for loss of information due to dropout when calculating sample size. Fourth, the study findings may not be generalizable to other populations since the study was primarily designed for African American/non-Hispanic Black populations.

Conclusion

This study clarifies the feasibility and preliminary efficacy of a culturally tailored DKD-focused lifestyle intervention in African Americans with type 2 diabetes and CKD in terms of clinical, knowledge, self-care, and behavior outcomes. Statistically significant changes in the clinical outcomes of

HbA_{1c}, total cholesterol, LDL, and eGFR were observed following the study intervention. In addition, there were statistically significant increases in CKD self-efficacy, CKD knowledge, exercise, and blood sugar testing. Based on the results of this study, a trial to determine the efficacy of this intervention would be feasible in African Americans with type 2 diabetes and CKD. The findings from this study will also serve as preliminary data to inform the design of a large-scale appropriately powered randomized controlled trial to examine the efficacy of a culturally tailored lifestyle intervention in African Americans with comorbid diabetes and CKD in terms of clinical, knowledge, self-care, and behavior outcomes.

Acknowledgments

Authors' Contributions

MNO and LEE designed the study. LEE analyzed the data. MNO drafted the manuscript. All authors were involved in critical revision of the manuscript content. The final manuscript was approved by all the authors. LEE and MNO are guarantors of this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Pre-post study intervention sessions and intervention content.
[DOCX File , 14 KB - formative_v6i3e34029_app1.docx]

Multimedia Appendix 2
Data collection schedule and measures for the pre-post study intervention.
[DOCX File , 16 KB - formative_v6i3e34029_app2.docx]

References

1. Tuttle KR, Bakris GL, Bilous RW, Chiang JL, de Boer IH, Goldstein-Fuchs J, et al. Diabetic Kidney Disease: A Report From an ADA Consensus Conference. American Journal of Kidney Diseases 2014 Oct;64(4):510-533 [FREE Full text] [doi: 10.1053/j.ajkd.2014.08.001] [Medline: 25257325]

2. Bikbov B, Purcell C, Levey A, Smith M, Abdoli A, Abebe M, et al. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet 2020 Feb 29;395(10225):709-733 [FREE Full text] [doi: 10.1016/S0140-6736(20)30045-3] [Medline: 32061315]

3. American Diabetes Association. 11. Microvascular Complications and Foot Care: Standards of Medical Care in Diabetes-2020. Diabetes Care 2020 Jan;43(Suppl 1):S135-S151. [doi: 10.2337/dc20-S011] [Medline: 31862754]

4. Saran R, Robinson B, Abbott KC, Agodoa LYC, Bragg-Gresham J, Balkrishnan R, et al. US Renal Data System 2018 Annual Data Report: Epidemiology of Kidney Disease in the United States. Am J Kidney Dis 2019 Mar;73(3 Suppl 1):A7-A8 [FREE Full text] [doi: 10.1053/j.ajkd.2019.01.001] [Medline: 30798791]

5. Carrero JJ, Grams ME, Sang Y, Ärnlöv J, Gasparini A, Matsushita K, et al. Albuminuria changes are associated with subsequent risk of end-stage renal disease and mortality. Kidney Int 2017 Jan;91(1):244-251 [FREE Full text] [doi: 10.1016/j.kint.2016.09.037] [Medline: 27927597]

6. Bowe B, Xie Y, Li T, Mokdad AH, Xian H, Yan Y, et al. Changes in the US Burden of Chronic Kidney Disease From 2002 to 2016: An Analysis of the Global Burden of Disease Study. JAMA Netw Open 2018 Nov 02;1(7):e184412 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.4412] [Medline: 30646390]

7. National Diabetes Statistics Report 2020. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf [accessed 2022-03-05]

8. Tarver-Carr ME, Powe NR, Eberhardt MS, LaVeist TA, Kington RS, Coresh J, et al. Excess risk of chronic kidney disease among African-American versus white subjects in the United States: a population-based study of potential explanatory factors. J Am Soc Nephrol 2002 Sep;13(9):2363-2370 [FREE Full text] [doi: 10.1097/01.asn.0000026493.18542.6a] [Medline: 12191981]

9. Davies MJ, D'Alessio DA, Fradkin J, Kernan WN, Mathieu C, Mingrone G, et al. Management of hyperglycaemia in type 2 diabetes, 2018. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetologia 2018 Dec;61(12):2461-2498. [doi: 10.1007/s00125-018-4729-5] [Medline: 30288571]

XSL•FO

RenderX

10. Perkovic V, Agarwal R, Fioretto P, Hemmelgarn BR, Levin A, Thomas MC, Conference Participants. Management of patients with diabetes and CKD: conclusions from a "Kidney Disease: Improving Global Outcomes" (KDIGO) Controversies Conference. Kidney Int 2016 Dec;90(6):1175-1183 [FREE Full text] [doi: 10.1016/j.kint.2016.09.010] [Medline: 27884312]

11. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney Int Suppl. 2013. URL: https://kdigo.org/wp-content/uploads/2017/02/KDIGO_2012_CKD_GL.pdf [accessed 2022-03-05]

12. Evangelidis N, Craig J, Bauman A, Manera K, Saglimbene V, Tong A. Lifestyle behaviour change for preventing the progression of chronic kidney disease: a systematic review. BMJ Open 2019 Oct 28;9(10):e031625 [FREE Full text] [doi: 10.1136/bmjopen-2019-031625] [Medline: 31662393]

13. Van Huffel L, Tomson CRV, Ruige J, Nistor I, Van Biesen W, Bolignano D. Dietary restriction and exercise for diabetic patients with chronic kidney disease: a systematic review. PLoS One 2014;9(11):e113667 [FREE Full text] [doi: 10.1371/journal.pone.0113667] [Medline: 25423489]

14. Look AHEAD Research Group, Pi-Sunyer X, Blackburn G, Brancati FL, Bray GA, Bright R, et al. Reduction in weight and cardiovascular disease risk factors in individuals with type 2 diabetes: one-year results of the look AHEAD trial. Diabetes Care 2007 Jun;30(6):1374-1383 [FREE Full text] [doi: 10.2337/dc07-0048] [Medline: 17363746]

15. Look AHEAD Research Group. Effect of a long-term behavioural weight loss intervention on nephropathy in overweight or obese adults with type 2 diabetes: a secondary analysis of the Look AHEAD randomised clinical trial. The Lancet Diabetes & Endocrinology 2014 Oct;2(10):801-809 [FREE Full text] [doi: 10.1016/S2213-8587(14)70156-1] [Medline: 25127483]

16. Ricardo AC, Anderson CA, Yang W, Zhang X, Fischer MJ, Dember LM, CRIC Study Investigators. Healthy lifestyle and risk of kidney disease progression, atherosclerotic events, and death in CKD: findings from the Chronic Renal Insufficiency Cohort (CRIC) Study. Am J Kidney Dis 2015 Mar;65(3):412-424 [FREE Full text] [doi: 10.1053/j.ajkd.2014.09.016] [Medline: 25458663]

17. Waterman AD, Browne T, Waterman BM, Gladstone EH, Hostetter T. Attitudes and behaviors of African Americans regarding early detection of kidney disease. Am J Kidney Dis 2008 Apr;51(4):554-562 [FREE Full text] [doi: 10.1053/j.ajkd.2007.12.020] [Medline: 18371531]

18. Umeukeje EM, Wild MG, Maripuri S, Davidson T, Rutherford M, Abdel-Kader K, et al. Black Americans' Perspectives of Barriers and Facilitators of Community Screening for Kidney Disease. Clin J Am Soc Nephrol 2018 Apr 06;13(4):551-559 [FREE Full text] [doi: 10.2215/CJN.07580717] [Medline: 29545381]

19. Afkarian M, Zelnick LR, Hall YN, Heagerty PJ, Tuttle K, Weiss NS, et al. Clinical Manifestations of Kidney Disease Among US Adults With Diabetes, 1988-2014. JAMA 2016 Aug 09;316(6):602-610 [FREE Full text] [doi: 10.1001/jama.2016.10924] [Medline: 27532915]

20. Williams JS, Lynch CP, Knapp RG, Egede LE. Technology-Intensified Diabetes Education Study (TIDES) in African Americans with type 2 diabetes: study protocol for a randomized controlled trial. Trials 2014 Nov 25;15:460 [FREE Full text] [doi: 10.1186/1745-6215-15-460] [Medline: 25425504]

21. Fisher JD, Fisher WA. The Information-Motivation-Behavioral Skills Model. In: DiClemente RJ, Crosby R, Kegler MC, editors. Emerging Theories in Health Promotion Practice and Research: Strategies for Improving Public Health. Indianapolis, IN: Jossey-Bass, Inc; 2002.

22. National Kidney Disease Education Program. National Institutes of Health. URL: https://www.niddk.nih.gov/health-information/community-health-outreach/information-clearinghouses/nkdep?dkrd=hispt1324 [accessed 2022-03-05]

23. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

24. Lin C, Wu C, Anderson RM, Chang C, Chang S, Hwang S, et al. The chronic kidney disease self-efficacy (CKD-SE) instrument: development and psychometric evaluation. Nephrol Dial Transplant 2012 Oct;27(10):3828-3834 [FREE Full text] [doi: 10.1093/ndt/gfr788] [Medline: 22344776]

25. Wright JA, Wallston KA, Elasy TA, Ikizler TA, Cavanaugh KL. Development and results of a kidney disease knowledge survey given to patients with CKD. Am J Kidney Dis 2011 Mar;57(3):387-395 [FREE Full text] [doi: 10.1053/j.ajkd.2010.09.018] [Medline: 21168943]

26. Garcia AA, Villagomez ET, Brown SA, Kouzekanani K, Hanis CL. The Starr County Diabetes Education Study: development of the Spanish-language diabetes knowledge questionnaire. Diabetes Care 2001 Jan;24(1):16-21. [doi: 10.2337/diacare.24.1.16] [Medline: 11194219]

27. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med 2004 Sep;36(8):588-594 [FREE Full text] [Medline: 15343421]

28. Toobert DJ, Hampson SE, Glasgow RE. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. Diabetes Care 2000 Jul;23(7):943-950. [doi: 10.2337/diacare.23.7.943] [Medline: 10895844]

29. Appel LJ, Wright JT, Greene T, Agodoa LY, Astor BC, Bakris GL, AASK Collaborative Research Group. Intensive blood-pressure control in hypertensive chronic kidney disease. N Engl J Med 2010 Sep 02;363(10):918-929 [FREE Full text] [doi: 10.1056/NEJMoa0910975] [Medline: 20818902]

30. de Zeeuw D, Ramjit D, Zhang Z, Ribeiro AB, Kurokawa K, Lash JP, et al. Renal risk and renoprotection among ethnic groups with type 2 diabetic nephropathy: a post hoc analysis of RENAAL. Kidney Int 2006 May;69(9):1675-1682 [FREE Full text] [doi: 10.1038/sj.ki.5000326] [Medline: 16572114]

31. Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein HC, Miller ME, Byington RP, Goff DC, Bigger JT, et al. Effects of intensive glucose lowering in type 2 diabetes. N Engl J Med 2008 Jun 12;358(24):2545-2559 [FREE Full text] [doi: 10.1056/NEJMoa0802743] [Medline: 18539917]

32. Rahman M, Pressel S, Davis BR, Nwachuku C, Wright JT, Whelton PK, et al. Renal outcomes in high-risk hypertensive patients treated with an angiotensin-converting enzyme inhibitor or a calcium channel blocker vs a diuretic: a report from the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). Arch Intern Med 2005 Apr 25;165(8):936-946. [doi: 10.1001/archinte.165.8.936] [Medline: 15851647]

33. ADVANCE Collaborative Group, Patel A, MacMahon S, Chalmers J, Neal B, Billot L, et al. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. N Engl J Med 2008 Jun 12;358(24):2560-2572. [doi: 10.1056/NEJMoa0802987] [Medline: 18539916]

34. Zimbudzi E, Lo C, Misso ML, Ranasinha S, Kerr PG, Teede HJ, et al. Effectiveness of self-management support interventions for people with comorbid diabetes and chronic kidney disease: a systematic review and meta-analysis. Syst Rev 2018 Jun 13;7(1):84 [FREE Full text] [doi: 10.1186/s13643-018-0748-z] [Medline: 29898785]

35. Tonneijck L, Muskiet MHA, Smits MM, van Bommel EJ, Heerspink HJL, van Raalte DH, et al. Glomerular Hyperfiltration in Diabetes: Mechanisms, Clinical Significance, and Treatment. J Am Soc Nephrol 2017 Apr;28(4):1023-1039 [FREE Full text] [doi: 10.1681/ASN.2016060666] [Medline: 28143897]

36. Zelniker TA, Wiviott SD, Raz I, Im K, Goodrich EL, Bonaca MP, et al. SGLT2 inhibitors for primary and secondary prevention of cardiovascular and renal outcomes in type 2 diabetes: a systematic review and meta-analysis of cardiovascular outcome trials. The Lancet 2019 Jan 05;393(10166):31-39 [FREE Full text] [doi: 10.1016/S0140-6736(18)32590-X] [Medline: 30424892]

37. Marso SP, Daniels GH, Brown-Frandsen K, Kristensen P, Mann JFE, Nauck MA, LEADER Steering Committee, LEADER Trial Investigators. Liraglutide and Cardiovascular Outcomes in Type 2 Diabetes. N Engl J Med 2016 Jul 28;375(4):311-322 [FREE Full text] [doi: 10.1056/NEJMoa1603827] [Medline: 27295427]

38. Marso SP, Bain SC, Consoli A, Eliaschewitz FG, Jódar E, Leiter LA, SUSTAIN-6 Investigators. Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. N Engl J Med 2016 Nov 10;375(19):1834-1844. [doi: 10.1056/NEJMoa1607141] [Medline: 27633186]

39. Hernandez AF, Green JB, Janmohamed S, D'Agostino RB, Granger CB, Jones NP, Harmony Outcomes committeesinvestigators. Albiglutide and cardiovascular outcomes in patients with type 2 diabetes and cardiovascular disease (Harmony Outcomes): a double-blind, randomised placebo-controlled trial. Lancet 2018 Oct 27;392(10157):1519-1529 [FREE Full text] [doi: 10.1016/S0140-6736(18)32261-X] [Medline: 30291013]

40. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, EMPA-REG OUTCOME Investigators. Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. N Engl J Med 2015 Nov 26;373(22):2117-2128. [doi: 10.1056/NEJMoa1504720] [Medline: 26378978]

41. Neal B, Perkovic V, Mahaffey KW, de Zeeuw D, Fulcher G, Erondu N, CANVAS Program Collaborative Group. Canagliflozin and Cardiovascular and Renal Events in Type 2 Diabetes. N Engl J Med 2017 Aug 17;377(7):644-657. [doi: 10.1056/NEJMoa1611925] [Medline: 28605608]

42. Wiviott SD, Raz I, Bonaca MP, Mosenzon O, Kato ET, Cahn A, DECLARE–TIMI 58 Investigators. Dapagliflozin and Cardiovascular Outcomes in Type 2 Diabetes. N Engl J Med 2019 Jan 24;380(4):347-357. [doi: 10.1056/NEJMoa1812389] [Medline: 30415602]

43. Holman RR, Bethel MA, Mentz RJ, Thompson VP, Lokhnygina Y, Buse JB, EXSCEL Study Group. Effects of Once-Weekly Exenatide on Cardiovascular Outcomes in Type 2 Diabetes. N Engl J Med 2017 Sep 28;377(13):1228-1239 [FREE Full text] [doi: 10.1056/NEJMoa1612917] [Medline: 28910237]

44. Pfeffer MA, Claggett B, Diaz R, Dickstein K, Gerstein HC, Køber LV, ELIXA Investigators. Lixisenatide in Patients with Type 2 Diabetes and Acute Coronary Syndrome. N Engl J Med 2015 Dec 03;373(23):2247-2257 [FREE Full text] [doi: 10.1056/NEJMoa1509225] [Medline: 26630143]

45. Harding K, Mersha TB, Vassalotti JA, Webb FA, Nicholas SB. Current State and Future Trends to Optimize the Care of Chronic Kidney Disease in African Americans. Am J Nephrol 2017;46(2):176-186 [FREE Full text] [doi: 10.1159/000479481] [Medline: 28787720]

46. Mitchell S, Bragg A, Moldovan I, Woods S, Melo K, Martin-Howard J, et al. Stigma as a Barrier to Participant Recruitment of Minority Populations in Diabetes Research: Development of a Community-Centered Recruitment Approach. JMIR Diabetes 2021 May 03;6(2):e26965 [FREE Full text] [doi: 10.2196/26965] [Medline: 33938811]

47. Wesson DE, Kitzman H, Montgomery A, Mamun A, Parnell W, Vilayvanh B, et al. A population health dietary intervention for African American adults with chronic kidney disease: The Fruit and Veggies for Kidney Health randomized study. Contemp Clin Trials Commun 2020 Mar;17:100540 [FREE Full text] [doi: 10.1016/j.conctc.2020.100540] [Medline: 32090186]

XSL•FO
RenderX

48.   Egede LE, Walker RJ. Structural Racism, Social Risk Factors, and Covid-19 - A Dangerous Convergence for Black Americans. N Engl J Med 2020 Sep 17;383(12):e77 [FREE Full text] [doi: 10.1056/NEJMp2023616] [Medline: 32706952]

Abbreviations

CKD: chronic kidney disease
DKD: diabetic kidney disease
eGFR: estimated glomerular filtration rate
HbA$_{1c}$: hemoglobin A$_{1c}$
LDL: low-density lipoprotein
PHQ-9: Patient Health Questionnaire-9
SDSCA: Summary of Diabetes Self-Care Activities
UACR: urine albumin to creatinine ratio

Original Paper

A Platform (Authorships.org) for the Objective Qualification and Order of Academic Authorship in Medical and Science Journals: Development and Evaluation Study Using the Design Science Research Methodology

Dimitri Aristotle Raptis[1,2,3], MSc, MD, PhD; Aristotle Raptis[4], BSc, MSc, PhD; Pascale Tinguely[1], MD; Authorships.org Collaborative[5]

[1]Department of Hepato-Pancreatico-Biliary Surgery and Liver Transplantation, Royal Free Hospital, London, United Kingdom

[2]Division of Surgical and Interventional Science, University College London, London, United Kingdom

[3]Department of Surgery and Transplantation, University Hospital Zurich, Zurich, Switzerland

[4]Division of Information and Communications Technology, University of Athens, Athens, Greece

[5]See Acknowledgments

Corresponding Author:
Dimitri Aristotle Raptis, MSc, MD, PhD
Department of Hepato-Pancreatico-Biliary Surgery and Liver Transplantation
Royal Free Hospital
Pond Street
London
United Kingdom
Phone: 44 020 7794 0500
Email: Dimitri.raptis@nhs.net

Abstract

Background: The qualification and order of authorship in scientific manuscripts are the main disputes in collaborative research work.

Objective: The aim of this project was to develop an open-access web-based platform for objective decision-making of authorship qualification and order in medical and science journals.

Methods: The design science process methodology was used to develop suitable software for authorship qualification and order. The first part of the software was designed to differentiate between qualification for authorship versus acknowledgment, using items of the recommendations of the International Committee of Medical Journal Editors. The second part addressed the order of authorship, using the analytical hierarchy process for objective multiple criteria decision-making and ranking. The platform was evaluated qualitatively (n=30) and quantitatively (n=18) using a dedicated questionnaire, by an international panel of medical and biomedical professionals and research collaborators worldwide.

Results: Authorships.org represents an open-access software compatible with all major platforms and web browsers. Software usability and output were evaluated and presented for 3 existing clinical and biomedical research studies. All 18 international evaluators felt that the Authorships.org platform was easy to use or remained neutral. Moreover, 59% (n=10) were satisfied with the software output results while the rest were unsure, 59% (n=10) would definitely use it for future projects while 41% (n=7) would consider it, 94% (n=16) felt it may prove useful to eliminate disputes regarding authorship, 82% (n=14) felt that it should become mandatory for manuscript submission to journals, and 53% (n=9) raised concerns regarding the potential unethical use of the software as a tool.

Conclusions: Authorships.org allows transparent evaluation of authorship qualification and order in academic medical and science journals. Objectified proof of authorship contributions may become mandatory during manuscript submission in high-quality academic journals.

XSL•FO
RenderX

KEYWORDS

authorship; writing; dissent and disputes; research ethics; software design

Introduction

Next to the presentation of scientific results, authorship in academic journal articles is a means for scholars to communicate the intellectual contributions of their work, take public responsibility for it, build reputation among peers, and convey their professional benefits [1]. However, scholars frequently encounter disputes concerning authorship, with the qualification and order of authorship remaining the main controversial issues in most collaborative work worldwide [2]. In other words, most dissents involve "Who are the authors, who should be acknowledged, and what should the order of the authors be in a given manuscript?" Numerous types of authorship abuses are considered scientific misconduct, with coercion (ie, intimidation tactics to gain authorship), guest, mutual support, ghost, and denial of authorship being the most frequent ones [3].

To address these issues, the International Committee of Medical Journal Editors (ICMJE) [4] established authorship guidelines including items related to study conception, execution, and documentation, which are adopted by many prominent institutions and journals worldwide [5]. While providing specific definitions regarding the roles of individual authors, the ICMJE guidelines do not allow a quantitative objectifiable assessment of individual author contributions within a body of authors, and disputes regarding authorship and the order of appearance frequently remain [6,7].

With the interest of promoting the highest ethics in medical and science publications, the aim of this project was to develop a user-friendly, open access, web-based software platform for the objective assessment of authorship qualification and order, in an attempt to reduce and hopefully eliminate authorship disputes.

Methods

Study Design

The design science research methodology was used to develop a suitable software for achieving our objective, as previously described [8]. Briefly, this is an established set of analytical techniques for performing research in information systems, involving the design of innovative products. The design science research methodology typically involves the problem identification, solution design, and evaluation phases [9].

Settings, Developers, and Collaborators

The software was designed by DAR located in Switzerland and developed together with AR located in Athens, Greece, starting on May 7, 2016. A group of collaborators worldwide contributed in the design, development, and evaluation of the software, and were selected based on personal contacts and through personal interest, with a full list of names available in the Acknowledgments and on Authorships.org. A subset of 18 of these collaborators were selected for formal software evaluation as described below, with eligibility for selection being a scientific background in medicine or other biomedical areas, regardless of the grade. The characteristics of the collaborators who formally evaluated the software can be found in Table 1. The overall software design and development duration was from May 2016 to December 2019. The software evaluation took place between March and October 2016, with the headquarters of software evaluation located in Zurich, Switzerland.

Table 1. Characteristics of the collaborators formally evaluating the software.

Gender	Job title	Academic degree	Speciality	Country
Male	Attending	MSc, MD, PhD	Surgery	Switzerland
Male	Research scientist	MD, PhD	Internal medicine	United States
Male	Assistant professor	BSc, PhD	Information technology	Switzerland
Male	Resident	MD	Surgery	Switzerland
Female	Resident	MD	Surgery	United Kingdom
Male	Resident	MD	Surgery	Switzerland
Male	Research scientist	BSc	Information technology	Switzerland
Male	Attending	MD	Surgery	Switzerland
Female	Resident	MD	Internal medicine	Switzerland
Male	Attending	MD	Internal medicine	Germany
Male	Attending	MD	Surgery	Switzerland
Male	Research scientist	MD	Surgery	Switzerland
Male	Attending	MD	Surgery	Switzerland
Male	Research scientist	MD	Basic science	Australia
Female	Resident	MD	Surgery	Switzerland
Male	Resident	BSc, Dipl Eng, MD	Internal medicine	Switzerland
Male	Resident	MD	Internal medicine	Switzerland
Male	Attending	MD	Surgery	Switzerland

Software Design and Development

The free web-based software is built on Drupal version 7 [10], an open-source content management system written in PHP and distributed under the GNU General Public License [11], as previously described [8]. An Apache server with a MySQL database was used [12]. The most advanced firewalls were installed for monitoring and prevention of malware. The costs for the server rental, domain name purchase, and software development were set at 2160 CHF (US $2332). These costs were successfully covered through a Kickstarter campaign by the collaborators on April 4, 2016 [13]. The website and software compatibility for different platforms, internet browsers, and devices were assessed using BrowserStack [14].

The first part of the software design focused on identifying and differentiating individuals whose work qualifies them as authors as opposed to contributors for acknowledgment. Items of the International Committee of "Medical Journal Editors Uniform Requirements for Manuscripts Submitted to Biomedical Journals," including conception, execution, and documentation of research, were used to design the first part of the software (Figure 1) [4].

The second part was designed to address the order of the qualifying authors using the analytical hierarchy process (AHP) method for objective multiple criteria decision-making and ranking [15]. Briefly, the AHP first decomposes the decision-making problem into a hierarchy of subproblems. Then, the relative weight of importance of the different criteria is assessed by pairwise comparisons. These weights are then used to calculate a score for each selection alternative. Information is decomposed into a hierarchy of alternatives, and criteria information is then synthesized to determine the relative ranking of alternatives. Both qualitative and quantitative information can be compared using informed judgements to derive weights and priorities. The consistency index measures the extent to which the decision-makers were consistent in their responses (Figure 2) [16].

Figure 1. Qualification for authorship versus acknowledgment.
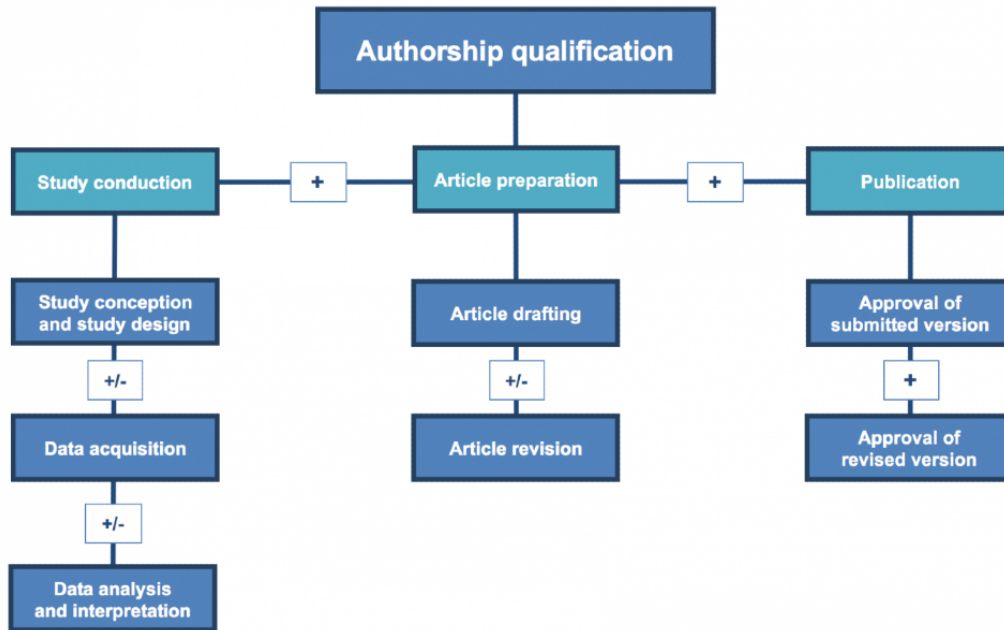
Figure 2. Analytical hierarchy process method for multiple comparisons and ranking of authors' contributions.
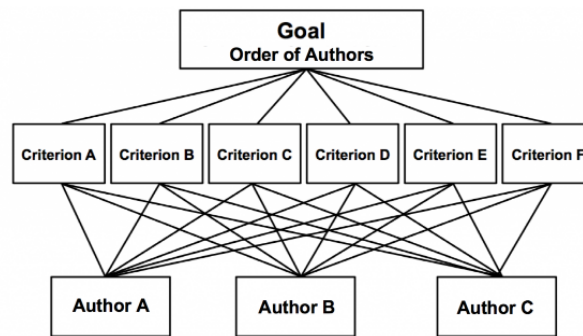
Software Evaluation

The software was evaluated and re-evaluated during the design phase as described above. At the end of software development, 18 collaborators (Table 1) assessed software usability and output in 3 existing clinical and biomedical research projects. The collaborators were chosen from the international panel of medical and biomedical professionals, based on the availability of a current research manuscript ready for submission and willingness to apply Authorships.org to evaluate their choice and order of authors. Collaborators were further asked to complete an online questionnaire (Multimedia Appendix 1). The evaluation questionnaire was designed by the authors, focusing on assessing the collaborator's attributes regarding authorship disputes, software functionality and usability, and the need for improvements [17]. Descriptive statistics were used to report results using rBiostatistics [18].

Results

Software Design

The developed software, named Authorships.org [19], is freely available online and is compatible with all major platforms, web browsers, and mobile or tablet devices [14]. Authorships.org performs server-side calculations and graphical rendering, which eliminates any hardware requirements or incompatibilities at the user side.

First, the user is requested to add the number and names of the individuals who participated in a particular research project. Next, their contribution to the work is assessed based on the generally agreed upon ICMJE criteria, such as conception, execution, documentation, and final approval for publication (Figure 3) [4]. The user is also given the choice to give equal or different weights to these criteria, in order to respect local norms at different institutions worldwide.

Based on the existing ICMJE guidelines, the software then indicates which individuals qualify for formal authorship versus for acknowledgment as a contributor. The user is subsequently asked to indicate a senior/last author of the manuscript, if required (Figure 4).

In the second part, the user assesses the extent of the individual contributions of all qualifying authors, except the one defined as senior/last author. A ranking of each author's extent of contribution to each ICMJE criterion is performed by using the AHP multiple comparisons method. Thereby, each author is compared to each other author in pairwise comparisons, using a decision matrix for ranking of the extent of contribution (Figure 5).

After all pairwise comparisons are made, a consistency ratio is provided to the user to ensure the input data are consistent and logical (Figure 6). This ratio should typically be below 0.10; however, values as high as 0.30 are acceptable, especially when large numbers of authors and their contributions are being assessed [15]. In case of severe inconsistency, the user is required to repeat the evaluation steps of the second part, ensuring comparisons are meaningful (eg, A > B, B > C, A > C).

Based on these quantitative and objective evaluations, the software defines the appropriate order of author appearance in a publication at a medical or science journal (Figure 7).

Figure 3. Selection of study contributions according to the International Committee of Medical Journal Editors (ICMJE) criteria [4].

Figure 4. Software output on qualification for authorship versus acknowledgement.

Figure 5. Example of pairwise comparison of the extent of individual author contributions for the International Committee of Medical Journal Editors (ICMJE) criterion of "study concept and design".
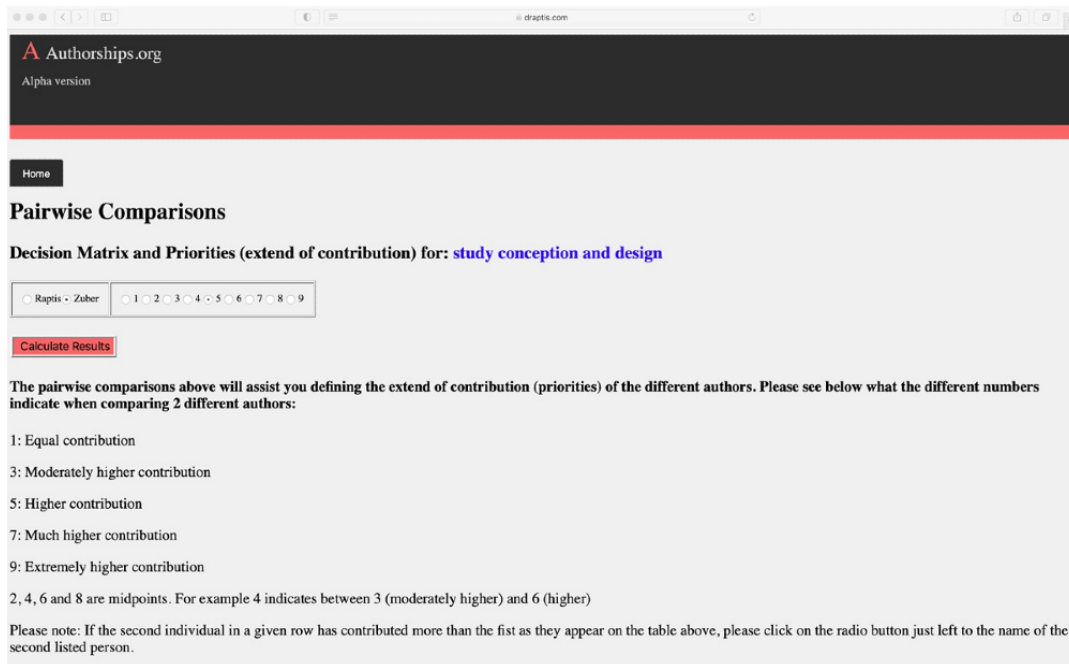
Figure 6. Software output on pairwise author comparisons and evaluation consistency.
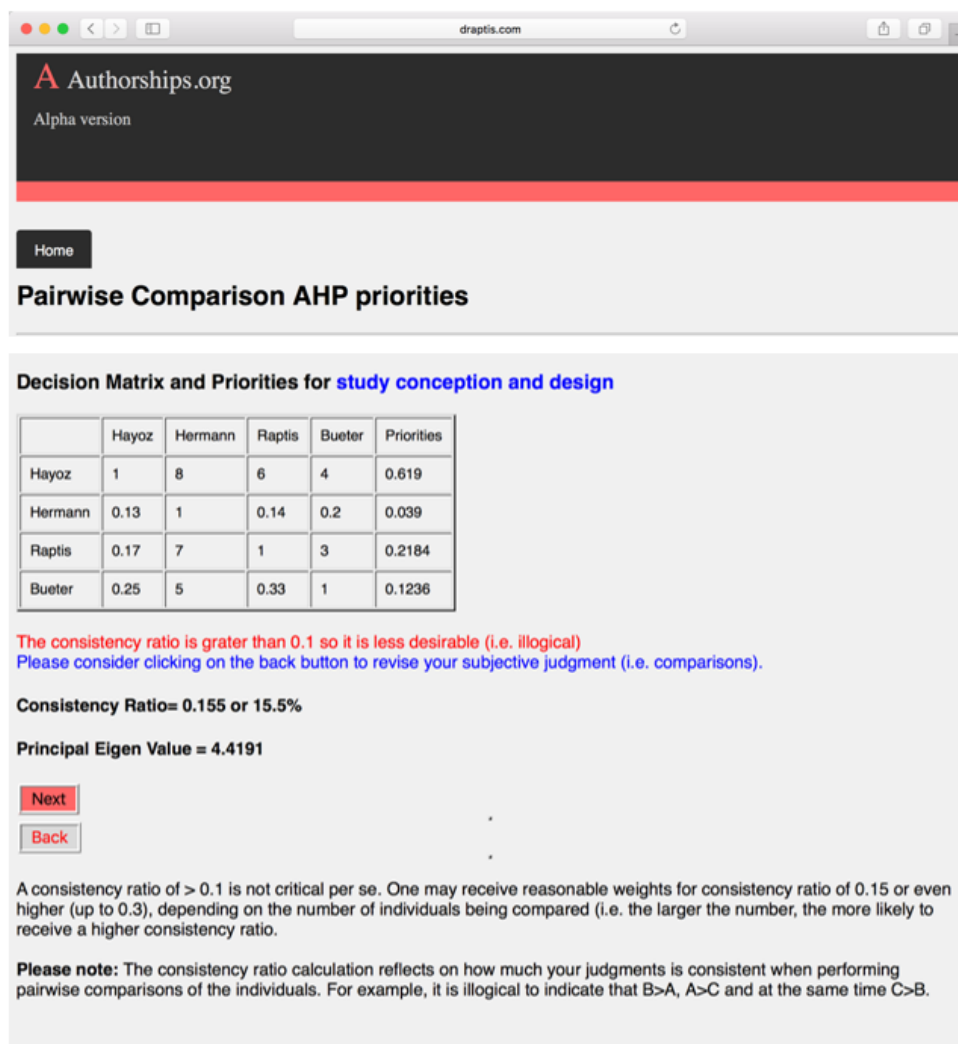
Figure 7. Software output on final author contribution ranking.
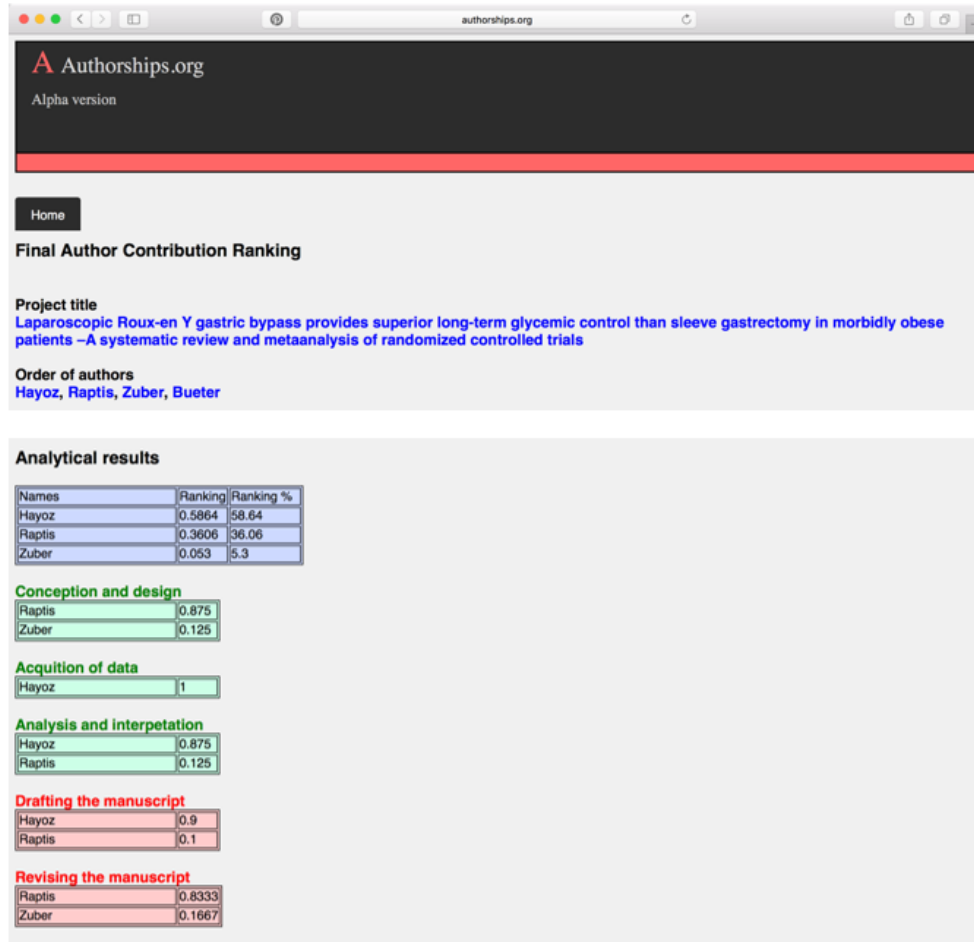
Software Evaluation

Questionnaire

Eighteen international collaborators evaluated the software using the specially designed questionnaire. Of these, 88% (n=15) felt that the software was easy to use, while 12% (n=3) remained neutral. Additionally, 59% (n=10) were satisfied with the software output results, feeling that it objectively reflects reality, while 35% (n=6) were unsure and 6% (n=1) were dissatisfied. The reason stated for dissatisfaction was that the collaborator felt "one very important contribution (ie, outstanding critical manuscript revision) may be sufficient to qualify for authorship;" however, authors with such contributions were disqualified when using Authorships.org. Fifty-nine percent (n=10) stated they would definitively use the software in future projects, while 41% (n=7) would potentially consider it. Moreover, 94% (n=16) felt that it may prove useful to eliminate disputes regarding authorship, and 82% (n=14) felt that it should become mandatory for manuscript submission to journals. Fifty-three percent (n=9) raised concerns regarding potential unethical use of the software as a tool for authorship evaluation. Reasons given were concerns that Authorships.org may represent one more tool to "justify" unethical authorship behavior. The quantitative results are summarized in Figure 8.

Figure 8. Results from the software evaluation questionnaire survey.
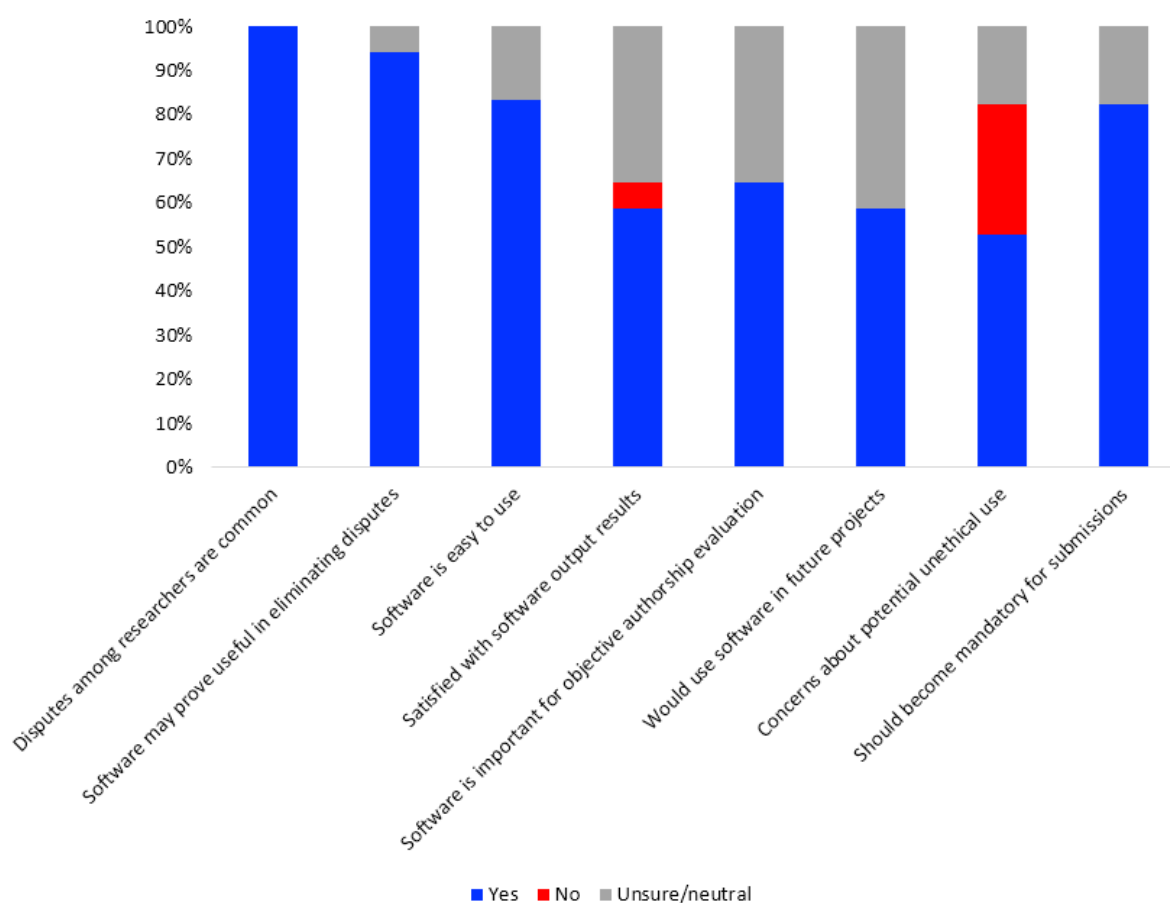
Case Study 1

In an attempt to objectify authorship disputes, 6 collaborators contributing to a clinical study were asked to provide the authors' order and reasoning behind it for a manuscript not yet submitted. The results of this survey were analyzed in a particular form developed at Authorships.org. Although they all agreed 100% regarding the authorship qualification of all 6 collaborators, a 0% total agreement on the order of author appearance was noted (Fleiss kappa: –0.10, 95% CI –0.18 to –0.01). In retrospect, the authors felt that the main reasons for this discordance were misconceptions on the amount of work conducted by their colleagues, lack of adherence to the "International Committee of Medical Journal Editors" recommendations, and some form of bias. Authorships.org was then applied to review the authorship order, yielding complete agreement of the collaborators. The collaborators felt that the platform was a transparent and objective tool for assigning authorship order, and justified the use of Authorships.org for their work.

Case Study 2

Seven authors of a biomedical engineering manuscript based on clinical and medical imaging data were asked to provide the order of the authors for the paper they all contributed. All authors agreed on their own qualification for authorship, and 2 (28%) authors agreed about the collaborator order for the manuscript. Authorships.org was used to assess author qualification and the order of authorship. Thereafter, 5 out of

the 7 authors justified the use of Authorships.org and thought that it is a robust and objective tool to assign the order of authors. The reason for the disagreement of the remaining 2 authors was a feeling that Authorships.org did not sufficiently acknowledge the seniority level of individual authors. They further disagreed to the design that authors should fulfil at least one green item, one red item, and the blue item to qualify for authorship in Autorship.org (based on the ICMJE recommendations) (Figure 3).

Case Study 3

Eight members of an international research group conducted a systematic review and meta-analysis regarding the use of comprehensive enhanced recovery protocols in the setting of liver transplantation. Individual tasks included study design and methodology, screening of titles and abstracts of studies identified in the literature search, data extraction and analysis of selected works, manuscript writing, and critical review. Authorships.org was applied after drafting the manuscript and before submission to a peer-reviewed scientific journal. The objective author qualification and order were applied directly in the manuscript as suggested by the software output and accepted by all authors.

Discussion

This study reports on the development and evaluation of a novel software tool for the objective assessment of authorship qualification and order, in an attempt to reduce authorship

disputes. To the best of our knowledge, Authorships.org represents the first open-access web-based software to quantitatively and thus objectively indicate the qualification and order of academic authorship in medical and science journals.

The majority of collaborators assessing the usability of Authorships.org found the software helpful and easy to use, and felt that the software could have a high impact in the scientific community in a short period of time. The concerns of the remaining collaborators regarding software output results focused on the criteria required for authorship qualification and therefore on the internationally accepted ICMJE criteria applied in the proposed software algorithm. While the qualification for authorship has been standardized in the ICMJE criteria [4], varying approaches of determining the order of authorship are currently applied across scientific disciplines, research groups, and countries. Examples of authorship policies include descending order of contribution, placing the person who took the lead in writing the manuscript or performing main research tasks first and the most experienced contributor last, or using an alphabetical or random order [1,20,21]. While the significance of a particular order may be understood in a given setting, the order of authorship is not generally agreed upon [22]. Attempts to reduce inaccuracies in author lists include models similar to the film credit concept, with mandatory contribution statements replacing the author's list, such as Credit Taxonomy [23]. However, these could lead to a shift in importance from the authors who actually produced the science to a more confounding way of giving credit to each research co-operator. Autorships.org not only represents a quantifiable and thus objectifiable approach, but also could provide a solution for the correct interpretation of the respective contributions of individual authors. Since the indication of specific author contributions has become mandatory in most scientific journals with a mid to high impact factor [22], the submission of a separate document objectifying the assessment of author qualification and contributions, such as provided by Authorships.org, may become obligatory in the future. To this end, Authorships.org must be formally validated in future projects worldwide, including a randomized controlled study comparing the levels of disputes.

One main hypothetical advantage obtained by the implementation of an objective instrument in academic authorship, as presented herein, is a reduction in the number of disputes. Academic authorship of papers arising from complex research projects, involving multiple centers and with potential valuable impact on the scientific community, can nowadays be considered as a currency for career development. These scenarios represent feeding terrains for conflicts between authors, which can sometimes even result in the retraction of a manuscript from publication or, if remaining unresolved due to failed mediation among authors, in rejection by the journal. Another known issue is the abuse of power exerted by certain authors inflicting the order of authors in their own interest, which can be avoided by using an objectifiable tool such as Authorships.org. We believe that the most reasonable attempt to resolve such disputes is undertaken primarily among the authors themselves, and in case of failure to compromise, the

disputes are taken up by the authors' affiliating institutions. Moreover, Authorships.org allows authors to evaluate the extent of contribution provided by each author to individual phases of manuscript preparation, representing the main advantage of using the AHP over other ranking models. Another main advantage of the proposed algorithm is the possibility to "customize" the relevance of each area of performance according to the contributors' previous agreement or the institution's internal guidelines. The use of the software could prompt a prearranged settlement and, consequently, help the team working on the project on the basis of mutually approved rules (eg, first authorship for a PhD candidate or last authorship for professorship title). This further allows maintaining the "motivating" aspect of a researcher to perform at her/his highest level when aiming to keep a certain position within the authorship list. Most of the collaborators evaluating the software further felt that the added consistency ratio supported them in reflecting their own contributions to a project in a quantitative way. A progressive improvement in consistency ratio values with increasing use of the software is further expected, as a result of ongoing education on the use of a mathematical way to evaluate each participant's contribution.

An important aspect of the Authorships.org design is the time point of the assessment of the qualification and order of authorship. While in many research groups worldwide, author roles are defined *prior to* the start of a research project, Authorships.org was deliberately designed to be applied *after* the execution and writing of a manuscript but before submission to a scientific journal, preferably in a discussion by all contributing members of the research team [1]. Hence, authorship qualification and order reflect the actual tasks performed by each individual rather than preset rigid orders potentially prone to abuse. It is essential to note that if not applied at the correct time point, Authorships.org may not be beneficial or, if abused, may even lead to erroneous results, which might represent a limitation to the current methodology. A further potential conflict of the proposed software is the option to change the weight of assessment for different parameters, which may increase the possibility of manipulation of weights toward personal interests. This was added to address the different needs across scientific fields. For example, in clinical research, more weight might be placed on the researcher who wrote the manuscript, whereas in basic science, more weight might be placed on the researcher performing the laboratory work. In general, we recommend giving equal weight to all assessment parameters. The acknowledgment of the seniority level of individual authors has been raised as a critical point by several collaborators evaluating Authorships.org. The software was designed such that 1 senior/last author could be specifically chosen based on her/his a priori qualification as the senior project leader. This was deliberately limited to 1 person to avoid multiple senior authors being listed for "political" reasons, acknowledging internal hierarchical structures rather actual contributions to a scientific work. While abusive authorship selections may still remain when using Authorships.org, the use of the proposed software has the potential to make such conflicts more transparent, raise awareness among authors, and therefore contribute to open discussions about conflicts in authorship contributions. Finally, views of the authorship disputes and

attributes, as evaluated by the collaborators, might not reflect generalizable views across researchers. However, with the selected collaborators originating from countries across the world and from varying areas of science, this potential selection bias might have been alleviated.

In conclusion, Authorships.org represents a novel approach to quantify and objectify the qualification and order of authorship in academic literature. It may become a mandatory tool for objectified proof of author contributions in scientific publications. Further randomized studies are needed to validate the potential of using such a tool for reducing or eliminating authorship disputes.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1
Online questionnaire.
[DOCX File , 79 KB - formative_v6i3e34258_app1.docx]

References

1. Credit where credit's due. Nature 2006 Mar 30;440(7084):591-592. [doi: 10.1038/440591a] [Medline: 16572137]
2. Dance A. Authorship: Who's on first? Nature 2012 Sep 27;489(7417):591-593 [FREE Full text] [doi: 10.1038/nj7417-591a] [Medline: 23025001]
3. Strange K. Authorship: why not just toss a coin? Am J Physiol Cell Physiol 2008 Sep;295(3):C567-C575 [FREE Full text] [doi: 10.1152/ajpcell.00208.2008] [Medline: 18776156]

XSL•FO
RenderX

4.   Defining the Role of Authors and Contributors. International Committee of Medical Journal Editors. URL: http://www. icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html [accessed 2021-08-15]

5.   Surgical Journal Editors Group. Consensus statement on surgery journal authorship--2006. Ann Surg 2006 Jun;243(6):713-714. [doi: 10.1097/00000658-200606000-00001] [Medline: 16772774]

6.   Macfarlane GJ. What to do about . . . authorship? Br J Pain 2021 Aug;15(3):249-250 [FREE Full text] [doi: 10.1177/20494637211023073] [Medline: 34377455]

7.   Hwang K. The Ajax Dilemma in Authorship. J Craniofac Surg 2021;32(5):1653-1654 [FREE Full text] [doi: 10.1097/SCS.0000000000007525] [Medline: 34319674]

8.   Raptis DA, Mettler T, Tzanas K, Graf R. A novel open-source web-based platform promoting collaboration of healthcare professionals and biostatisticians: a design science approach. Inform Health Soc Care 2012 Jan;37(1):22-36. [doi: 10.3109/17538157.2011.590257] [Medline: 22010822]

9.   Offermann P, Levina O, Schönherr M, Bub U. Outline of a design science research process. In: DESRIST '09: Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology. 2009 Presented at: 4th International Conference on Design Science Research in Information Systems and Technology; May 7-8, 2009; Philadelphia, PA p. 1-11. [doi: 10.1145/1555619.1555629]

10. Drupal 7. Drupal. URL: https://www.drupal.org/about/drupal-7 [accessed 2019-12-10]

11. GNU General Public License. GNU. URL: https://www.gnu.org/licenses/gpl-3.0.en.html [accessed 2019-12-20]

12. MySQL. URL: https://www.mysql.com [accessed 2019-11-15]

13. Software for ethical authorship in medical journals. Kickstarter. URL: https://www.kickstarter.com/projects/ebiostatistics/ software-for-ethical-authorship-in-medical-journal [accessed 2016-04-04]

14. BrowserStack. URL: https://www.browserstack.com [accessed 2018-10-30]

15. Saaty TL. The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. New York, NY: McGraw-Hill International Book Company; 1980.

16. Marinoni O. Implementation of the analytical hierarchy process with VBA in ArcGIS. Computers & Geosciences 2004 Jul;30(6):637-646 [FREE Full text] [doi: 10.1016/j.cageo.2004.03.010]

17. Folmer E, van Gurp J, Bosch J. Software Architecture Analysis of Usability. In: Bastide R, Palanque P, Roth J, editors. Engineering Human Computer Interaction and Interactive Systems. EHCI 2004. Lecture Notes in Computer Science, vol 3425. Berlin, Heidelberg: Springer; 2005:38-58.

18. rBiostatistics. URL: https://rbiostatistics.com [accessed 2021-08-01]

19. Authorships.org. URL: https://authorships.org/ [accessed 2021-08-10]

20. Waltman L. An empirical analysis of the use of alphabetical authorship in scientific publishing. Journal of Informetrics 2012 Oct;6(4):700-711 [FREE Full text] [doi: 10.1016/j.joi.2012.07.008]

21. Landry A, Zhao J, Shappell E. Authorship anatomy: a guide for scholars. BMJ 2020 Dec 14;371:m4702 [FREE Full text] [doi: 10.1136/bmj.m4702] [Medline: 33318053]

22. Yank V, Rennie D. Disclosure of researcher contributions: a study of original research articles in The Lancet. Ann Intern Med 1999 Apr 20;130(8):661-670. [doi: 10.7326/0003-4819-130-8-199904200-00013] [Medline: 10215563]

23. CRediT. URL: https://credit.niso.org [accessed 2021-08-10]

Abbreviations

AHP: analytical hierarchy process
ICMJE: International Committee of Medical Journal Editors

XSL•FO
RenderX

<u>Original Paper</u>

Psychometric Evaluation of a Fear of COVID-19 Scale in China: Cross-sectional Study

Edmond P H Choi[1], PhD; Wenjie Duan[2], PhD; Daniel Y T Fong[1], PhD; Kris Y W Lok[1], PhD; Mandy Ho[1], PhD; Janet Y H Wong[1], PhD; Chia-Chin Lin[1,3], PhD

[1]School of Nursing, The University of Hong Kong, Hong Kong, China (Hong Kong)

[2]Social and Public Administration School, East China University of Science and Technology, Shanghai, China

[3]School of Nursing, College of Nursing, Taipei Medical University, Taipei, Taiwan

Corresponding Author:
Daniel Y T Fong, PhD
School of Nursing
The University of Hong Kong
21 Sassoon Road, Pokfulam
Hong Kong
China (Hong Kong)
Phone: 852 39176645
Fax: 852 28726079
Email: dytfong@hku.hk

Abstract

Background: At the very beginning of the COVID-19 pandemic, information about fear of COVID-19 was very limited in Chinese populations, and there was no standardized and validated scale to measure the fear associated with the pandemic.

Objective: This cross-sectional study aimed to adapt and validate a fear scale to determine the levels of fear of COVID-19 among the general population in mainland China and Hong Kong.

Methods: A web-based questionnaire platform was developed for data collection; the study instruments were an adapted version of the 8-item Breast Cancer Fear Scale ("Fear Scale") and the 4-item Patient Health Questionnaire. The internal construct validity, convergent validity, known group validity, and reliability of the adapted Fear Scale were assessed, and descriptive statistics were used to summarize the participants' fear levels.

Results: A total of 2822 study participants aged 18 years or older were included in the analysis. The reliability of the adapted scale was satisfactory, with a Cronbach α coefficient of .93. The item-total correlations corrected for overlap were >0.4, confirming their internal construct validity. Regarding convergent validity, a small-to-moderate correlation between the Fear Scale and the 4-item Patient Health Questionnaire scores was found. Regarding known group validity, we found that the study participants who were recruited from Hong Kong had a higher level of fear than the study participants from mainland China. Older adults had a higher level of fear compared with younger adults. Furthermore, having hypertension, liver disease, heart disease, cancer, anxiety, and insomnia were associated with a higher fear level. The descriptive analysis found that more than 40% of the study participants reported that the thought of COVID-19 scared them. About one-third of the study participants reported that when they thought about COVID-19, they felt nervous, uneasy, and depressed.

Conclusions: The psychometric properties of the adapted Fear Scale are acceptable to measure the fear of COVID-19 among Chinese people. Our study stresses the need for more psychosocial support and care to help this population cope with their fears during the pandemic.

Introduction

In December 2019, the novel coronavirus disease 2019 (COVID-19) emerged in Wuhan City, China [1]. The outbreak rapidly evolved into a global pandemic [2], affecting more than 190 countries and regions [3]. The COVID-19 pandemic continues to spread on a global scale. As of December 20, 2021, there have been more than 271 million confirmed cases of COVID-19 worldwide, with more than 5 million deaths [4]. The COVID-19 pandemic has lasted for almost 2 years and is still ongoing. With time, an increasing number of COVID-19 variants was reported globally [5].

COVID-19 is not only life-threatening, but it also leads to psychological distress [6-10]. The concomitant public health measures such as quarantines, social distancing, and lockdowns can also increase psychosocial distress. Since the start of the COVID-19 pandemic, a plethora of research studies have been conducted to examine the psychological status of people during the pandemic. A meta-analysis of 55 peer-reviewed studies found that the prevalence of depression was 16%, the prevalence of anxiety was 15%, the prevalence of insomnia was 24%, and the prevalence of posttraumatic stress disorder was 22% [11]. Another meta-analysis of the prevalence of stress, anxiety, and depression among the general population during the COVID-19 pandemic found that the prevalence of stress was 29.6%, that of anxiety was 32%, and that of depression 34% [12]. Moreover, compared with studies conducted in Europe, those conducted in Asia found a higher prevalence of anxiety (Asia 33% vs Europe 24%) and depression (Asia 35% vs Europe 32%) [12]. These studies suggest that the pandemic substantially jeopardizes the psychological well-being of the general population [11,12].

In addition to anxiety, depression, and stress, fear is also a common psychological response to COVID-19 [13]. In brief, fear is an adaptive emotion that helps defend against potential danger [14]. Fear may occur in response to specific stimuli in the present environment or in anticipation of future or imagined situations that pose a threat to oneself [15]. During the COVID-19 pandemic, people may experience the fear of contracting the infection and a feeling of uncertainty. Fear can be beneficial because it can motivate people to engage in preventive behaviors, such as hand hygiene and mask wearing [16]. However, excessive fear can be maladaptive, leading to psychological distress. For example, fear of COVID-19 may exacerbate preexisting mental health and psychiatric conditions [17]. In extreme situations, fear may lead to suicidal ideation [18]. Excessive fear can cause irrational behaviors, such as panic buying [19].

It is noteworthy that the COVID-19 pandemic has reignited the fear resulting from the 2003 severe acute respiratory syndrome outbreak for many people in mainland China and Hong Kong. This adverse experience was unique to those populations. The fear levels of people in mainland China and Hong Kong may therefore be different from those in populations that did not undergo that adverse experience. Assessing and managing fear is a crucial component of outbreak control and health promotion [20].

In this study, the 8-item Breast Cancer Fear Scale developed by Champion et al [21] was used. We chose this instrument to measure fear levels for several reasons. First, at the very beginning of the pandemic, there was no standardized and validated study instrument specifically developed to measure fear levels related to COVID-19. For example, the Fear of COVID-19 Scale developed by Ahorsu and colleagues [13] was not available when we planned this study. Second, the 8-item Breast Cancer Fear Scale ("Fear Scale") was one of the few instruments available to measure fear among the Hong Kong Chinese population [22]. Furthermore, even though the Fear Scale was originally developed to measure fear related to breast cancer, the question items are generic and comprehensive. The Fear Scale covers common responses to fear such as feeling scared, nervous, upset, depressed, jittery, uneasy, and anxious, as well as having heart palpitations. According to a study in Canada, many participants felt uneasy, distressed, anxious, and nervous due to the COVID-19 pandemic [23]. A study in Slovakia reported an overall increase in negative feeling such as feeling upset, scared, and afraid during the COVID-19 pandemic [24,25]. The items of the Fear Scale should be applicable and appropriate to measure the fear related to COVID-19.

This study aimed to adapt and validate the Fear Scale to determine the levels of fear of COVID-19 in mainland China and Hong Kong. With the information on how an individual fears COVID-19, health care providers can design appropriate psychosocial interventions to meet the public's needs.

Methods

Study Design, Participants, and Sampling

An international study was conducted, which aimed to examine the global impact of the COVID-19 pandemic on lifestyle behaviors, fear, depression, and perceived needs of communities [26,27]. The study was conducted in 30 countries across the globe. It is a cross-sectional web-based survey design. Moreover, a web-based questionnaire platform was developed for data collection [28].

For this analysis, only data collected in mainland China and Hong Kong between July 2020 and January 2021 were used. Study eligibility criteria included (1) aged ≥18 years; (2) being able to read and understand Chinese; and (3) having an internet access. To recruit more people with diverse sociodemographic backgrounds, multiple recruitment strategies were used to recruit study participants. The study participants were recruited by survey service providers, social media platforms such as Facebook, WeChat, and Twitter, and snowball sampling, in which the existing study participants helped to recruit additional participants to join this study. To encourage more people to complete the survey, for each completed questionnaire, HK $1 (US $0.13) would be donated to the Red Cross in the respondent's region.

The study protocol has been published elsewhere [26]. The study was approved by the institutional review board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (reference UW 20-272). All the procedures involving

human participants in this study were conducted in accordance with the ethical standards of the institutional review board and the 1964 Declaration of Helsinki and its later amendments. Electronic informed consent was obtained from each study participant.

Outcomes and Instruments

The primary outcome of the study was the fear of COVID-19. To measure the fear of COVID-19, we adapted the 8-item Breast Cancer Fear Scale developed by Champion et al [21] for this study. The study instrument was originally developed to measure women's emotional responses to breast cancer. In the scale developed by Champion et al [21], a 5-point Likert scale is used (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, and 5=strongly agree); a higher score indicates a higher level of fear. The total score of the instrument is the sum of each item. In this paper, we changed the words "breast cancer" to "COVID-19" in all of the following 8 items: (1) the thought of COVID-19 scares me; (2) when I think about COVID-19, I feel nervous; (3) when I think about COVID-19, I get upset; (4) when I think about COVID-19, I get depressed; (5) when I think about COVID-19, I get jittery; (6) when I think about COVID-19, my heart beats faster; (7) when I think about COVID-19, I feel uneasy; and (8) when I think about COVID-19, I feel anxious.

The face validity of the adapted instrument was evaluated by an expert panel of this study.

The 4-item Patient Health Questionnaire (PHQ-4), which measures anxiety and depressive symptoms, was administered to evaluate the convergent validity of the Fear Scale. The PHQ-4 includes the 2-item Generalized Anxiety Disorder scale and the 2-item Patient Health Questionnaire (PHQ-2). A 4-point Likert scale is used (0=not at all; 1=several days; 2=more than half the days; and 3=nearly every day). The summary score of the PHQ-4 ranges from 0 to 12, with a higher score indicating greater anxiety and depressive symptoms. The PHQ-4 was validated in Chinese adults [29]. The study supported its 2-factor model and reliability [29]. Cronbach α coefficient was .87 for the PHQ-4, .80 for the 2-item Generalized Anxiety Disorder, and .80 for the PHQ-2 in this study.

A structured questionnaire was used to collect sociodemographic factors such as age, gender, and comorbidities.

Data Analysis

The internal construct validity, convergent validity, known group validity, and reliability of the Fear Scale were assessed. The internal construct validity was evaluated using the corrected item-total correlation; a correlation coefficient of ≥0.4 indicated adequate internal construct validity. The convergent validity of the Fear Scale was determined by calculating the Pearson correlation coefficient between the total score of the Fear Scale and the total score of the PHQ-4. It was hypothesized that an absolute value Pearson correlation coefficient of at least 0.3 was required [30]. To evaluate the known group validity, independent t tests were used to compare the mean score of the Fear Scale between (1) people recruited from mainland China and people recruited from Hong Kong [31]; (2) people aged 18-59 years and people aged 60 years or older [15]; and (3) male and female participants [32]. A study among Chinese university students reported that students in mainland China had lower fear of instability related to the COVID-19 pandemic when compared with students in Hong Kong [31]. Another study among pregnant women and new mothers reported that compared with the study participants in mainland China, the level of fear related to the COVID-19 pandemic was significantly higher among study participants in Hong Kong [33]. A study in Singapore found that older age was associated with greater fear of COVID-19 [15]. Another study in the Spanish population found that fear was higher among women than among men [32]. Besides, a study in Turkey reported that the COVID-19 fear scores were higher among people with a chronic disease [34]. Therefore, we also compared the mean score of the Fear Scale between people with and without the following chronic diseases, which were highly prevalent in Chinese populations: (1) hypertension; (2) diabetes; (3) liver disease; (4) heart disease; (5) stroke; (6) chronic obstructive pulmonary disease; (7) cancer; (8) depression; (9) anxiety; and (10) insomnia. Cohen d effect sizes were also calculated. The interpretation of the effect sizes was as follows: trivial (<0.2), small (≥0.2 to <0.5), moderate (≥0.5 to <0.8) and large (≥0.8). Finally, descriptive statistics were used to describe the fear levels of the study participants. Furthermore, multiple linear regression analysis was used to explore the known associations between sociodemographic and clinical factors, on the one hand, and the Fear Scale, on the other.

Results

Participants' Characteristics

A total of 2822 study participants aged 18 years or older were included in the analysis, of whom 61% (n=1721) were female and 38.8% (n=1096) were male. Three-quarters, 75.6% (n=2133), were recruited from Hong Kong while 24.4% (n=689) were recruited from mainland China. Almost half (1450, 51.4%) of the participants were married. Hypertension (294, 10.4%) was the most prevalent reported chronic disease, followed by diabetes (155, 5.5%), insomnia (90, 3.2%) and heart disease (62, 2.2%). Table 1 shows the participants' characteristics of the study.

Table 1. Sociodemographic and clinical profile (N=2822).

Characteristics	Values, n (%)
Geographical areas	
China	689 (24.4)
Hong Kong	2133 (75.6)
Age groups (years)	
18-59	2547 (90.3)
≥60	275 (9.7)
Genders	
Female	1721 (61)
Male	1096 (38.8)
Nonbinary	5 (0.2)
Marital status	
Single	1265 (44.8)
Married or cohabitation or common-law	1450 (51.4)
Separated or divorced or widowed	107 (3.8)
Educational level	
Primary or below	169 (6)
Secondary	1191 (42.2)
College	289 (10.2)
Associate degree	204 (7.2)
Bachelor's degree	780 (27.6)
Master's degree or above	189 (6.7)
Employment status	
Nonworking	589 (20.9)
Working	1682 (59.6)
Student	551 (19.5)
Chronic diseases	
Hypertension	294 (10.4)
Diabetes	155 (5.5)
Liver disease	50 (1.8)
Heart disease	62 (2.2)
Stroke	18 (0.6)
Chronic obstructive pulmonary disease	36 (1.3)
Cancer	16 (0.6)
Insomnia	90 (3.2)
Depression	43 (1.5)
Anxiety	44 (1.6)

Reliability and Validity of the Fear Scale

The mean score of the Fear Scale was 23.60 (SD 6.64), and the Cronbach α coefficient was .93. The corrected item-total correlations were >0.7 for all items. Table 2 shows the results of the internal consistency and internal construct validity. The Pearson correlation coefficient between the Fear Scale and PHQ-4 scores was 0.23 (P<.001). Table 3 shows the results of the convergent validity.

Table 2. Descriptive statistics, internal construct validity, and reliability of the Fear Scale (N=2822).

	Corrected item-total correlation (n=2821)	Mean (SD)[a]	Strongly disagree, n (%)	Disagree, n (%)	Neutral, n (%)	Agree, n (%)	Strongly agree, n (%)	Missing, n (%)
The thought of COVID-19 scares me	0.71	3.27 (0.98)	151 (5.4)	451 (16)	890 (31.5)	1145 (40.6)	185 (6.6)	N/A[b]
When I think about COVID-19, I feel nervous	0.73	3.11 (0.99)	187 (6.6)	527 (18.7)	1076 (38.1)	862 (30.6)	170 (6)	N/A
When I think about COVID-19, I get upset	0.75	2.99 (1.02)	225 (8)	642 (22.8)	1083 (38.4)	694 (24.6)	178 (6.3)	N/A
When I think about COVID-19, I get depressed	0.80	2.99 (1.03)	235 (8.3)	656 (23.3)	1007 (35.7)	760 (26.9)	163 (5.8)	1 (0.04)
When I think about COVID-19, I get jittery	0.71	2.63 (1.04)	449 (15.9)	810 (28.7)	992 (35.2)	480 (17)	91 (3.2)	N/A
When I think about COVID-19, my heart beats faster	0.70	2.64 (1.04)	426 (15.1)	823 (29.2)	1006 (35.6)	463 (16.4)	104 (3.7)	N/A
When I think about COVID-19, I feel uneasy	0.78	3.03 (1.04)	256 (9.1)	560 (19.8)	1003 (35.5)	854 (30.3)	149 (5.3)	N/A
When I think about COVID-19, I feel anxious	0.80	2.95 (1.03)	279 (9.9)	585 (20.7)	1101 (39)	719 (25.5)	138 (4.9)	N/A
Cronbach alpha	0.93	N/A	N/A	N/A	N/A	N/A	N/A	N/A

[a]A higher score means a higher level of fear.

[b]N/A: not applicable.

Table 3. Convergent validity of the Fear Scale (N=2822).

Scale	Population, n	Mean (SD)[a]	Pearson correlation coefficient	P value
PHQ-4[b] anxiety subscale or GAD-2[c]	2822	1.22 (1.26)	—[d]	—
PHQ-4 depression subscale or PHQ-2[e]	2821	1.10 (1.30)	—	—
PHQ-4 total score	2821	2.32 (2.37)	—	—
Fear Scale total score	2821	23.60 (6.64)	—	—
PHQ-4 anxiety subscale or GAD-2	2821	—	0.25	<.001
PHQ-4 depression subscale PHQ-2	2820	—	0.19	<.001
PHQ-4 total score	2820	—	0.23	<.001

[a]A higher score means a higher level of fear or anxiety or depression.

[b]PHQ-4: 4-item Patient Health Questionnaire.

[c]GAD-2: two-item Generalized Anxiety Disorder scale.

[d]Not available.

[e]PHQ-2: 2-item Patient Health Questionnaire.

With respect to the known group comparisons, the results of the independent t tests showed that the study participants who were recruited from Hong Kong had a higher level of fear compared with the study participants from mainland China (Cohen d effect size 0.24). Furthermore, older adults (60 years or above) had a higher level of fear than younger adults (Cohen d effect size 0.39). Study participants with cancer (Cohen d effect size 0.58), heart disease (Cohen d effect size 0.44), hypertension (Cohen d effect size 0.36), liver disease (Cohen d effect size 0.33), insomnia (Cohen d effect size 0.33), and anxiety (Cohen d effect size 0.28) had a higher level of fear than those without such conditions. Table 4 and Table 5 show the results of the known group comparisons by independent t test.

The results of multiple linear regression are shown in the Multimedia Appendix 1.

Table 4. Known group comparison: sociodemographic factors (N=2822).

Sociodemographic factors	n (%)	Fear Scale, mean (SD)	P value	Cohen d effect size
Location			<.001	0.24
Mainland China	688 (24.4)	22.40 (6.49)		
Hong Kong	2133 (75.6)	23.98 (6.64)		
Gender			.80	0.01
Female	1720 (60.9)	23.58 (6.60)		
Male	1096 (38.8)	23.64 (6.69)		
Age (years)			<.001	0.39
18-59	2546 (90.2)	23.34 (6.56)		
≥60	275 (9.7)	25.98 (6.86)		

Table 5. Known group comparison: clinical factors (N=2822).

Clinical factors	n (%)	Fear Scale, mean (SD)	P value	Cohen d effect size
Hypertension			<.001	0.36
No	2527 (89.5)	23.35 (6.59)		
Yes	294 (10.4)	25.73 (6.68)		
Diabetes			.11	0.13
No	2666 (94.5)	23.55 (6.63)		
Yes	155 (5.5)	24.43 (6.77)		
Liver disease			.02	0.33
No	2771 (98.2)	23.56 (6.63)		
Yes	50 (1.8)	25.78 (6.96)		
Heart disease			<.001	0.44
No	2759 (97.8)	23.53 (6.61)		
Yes	62 (2.2)	26.61 (7.27)		
Stroke			.16	0.31
No	2803 (99.3)	23.58 (6.63)		
Yes	18 (0.6)	25.78 (7.40)		
Chronic obstructive pulmonary disease			.28	0.19
No	2785 (98.7)	23.58 (6.65)		
Yes	36 (1.3)	24.78 (5.83)		
Cancer			.04	0.58
No	2805 (99.4)	23.58 (6.64)		
Yes	16 (0.6)	26.94 (4.74)		
Depression			.17	0.20
No	2778 (98.4)	23.58 (6.63)		
Yes	43 (1.5)	24.98 (7.40)		
Anxiety			.04	0.28
No	2777 (98.4)	23.57 (6.61)		
Yes	44 (1.6)	25.61 (8.01)		
Insomnia			.002	0.33
No	2731 (96.8)	23.53 (6.63)		
Yes	90 (3.2)	25.70 (6.47)		

The Fear Levels of the Study Participants

In total, 47.1% (n=1330) of the participants reported that the thought of COVID-19 scared them. Moreover, 36.6% (n=1032) of the study participants reported that they felt nervous when they thought about COVID-19. About one-third of the participants reported that they felt uneasy (1003, 35.5%) and became depressed (923, 32.7%) when they thought about COVID-19. The descriptive statistics of the Fear Scale are shown in Table 2.

We also separated the analysis between data collected in Hong Kong and those collected in China. Those results are shown in Multimedia Appendix 1.

Discussion

Principal Results

In the first part of this study, we assessed the psychometric properties in terms of internal construct validity, convergent validity, known group validity, and reliability of the Fear Scale. The Cronbach α coefficient was .93, which is far larger than the recommended cut-off value of .7. This finding supports the general agreement between the 8 items that make up the composite score of the scale to measure the fear related to COVID-19. Moreover, we found that all the coefficients of the item-total correlation, corrected for overlaps, were larger than 0.4, supporting the internal construct validity of the modified scale. These results supported the suggestion that all individual items measured the same construct as that measured by the other items. With regards to the study's convergent validity, we found a small-to-moderate correlation between the total score of the Fear Scale and the total score of the PHQ-4. Another important finding of this study was that participants with a chronic disease had a higher fear level than those without a chronic disease. Particularly, we found that hypertension, liver disease, heart disease, cancer, anxiety, and insomnia were associated with a higher fear level.

Limitations

There were some limitations in this study. First, the study was conducted in mainland China and Hong Kong. Therefore, the study findings may not be transferable to other geographic areas in which the severity of COVID-19, case fatality rate, and infection control measures are different. We expect that the fear level would be even higher in areas where the severity and case fatality rate of COVID-19 were more severe. Second, we could not explore the trajectory of the fear levels over time due to the cross-sectional nature of the study. Third, we adapted the Breast Cancer Fear Scale in this study; thus, some of the constructs related to COVID-19 could not be measured. However, as previously mentioned, there was no validated fear scale specific to COVID-19 when we planned our study. Fourth, regarding the reliability of the scale, we only evaluated its internal consistency. We were not able to evaluate the test-retest reliability of the scale due to the cross-sectional design of the study. Fifth, regarding the known group comparison, the sample size of some subgroups was small such as that of patients with diabetes and depression. There might be insufficient statistical power to detect the differences between groups. Finally, we

used a web-based questionnaire platform to collect the data. People with low computer literacy would probably be excluded from the study. Accordingly, the potential sampling bias should be noted.

Comparisons With Prior Work

We found that the total score of the Fear Scale had a higher correlation with the PHQ-4 anxiety subscale than with the PHQ-4 depression subscale. In fact, there are distinct differences in psychological features between fear and depression. According to Witte [35], fear is conceptualized as negatively toned emotion accompanied by a high level of physiologic arousal stimulated by a threat. Fear can be expressed as a physiological arousal, such as feeling "jittery" and "heart beating faster," through verbal self-reports of fear (eg, "I feel scared") and overt acts that exhibit fear, such as facial expression [21]. These emotional and physiological reactions to perceived threats are fundamentally different from those of depression, which is manifested through the following 4 symptom clusters: (1) emotional symptoms such as feeling sad and worthless; (2) cognitive symptoms such as a negative view of the self and hopelessness; (3) motivational symptoms such as a lack of incentive; and (4) somatic symptoms such as a loss of appetite and sleep disturbances [36,37]. Additionally, it was suggested that fear and anxiety are largely distinct emotions. A meta-analysis reported only a moderate (r=0.32) relationship between measures of trait fear and anxiety [38]. Fear is an aversive psychological state during which an individual is motivated to escape a specific and imminent threat. The characteristics of fear include short-lived arousal that quickly dissipates after the threat is avoided. By contrast, anxiety is an aversive psychological state that occurs while an individual approaches an ambiguous and uncertain threat. Hypervigilance and hyperarousal are the typical behaviors characteristic of anxiety [38]. The small-to-moderate correlation between the Fear Scale and the PHQ-4 further supported the need for this study, which adapted and validated the Fear Scale to measure the fear of COVID-19. Besides, compared with study participants recruited in Hong Kong, those recruited from mainland China had a higher PHQ-4 score but lower Fear Scale total score. This finding further suggested that the constructs of fear and anxiety are different.

Participants with a chronic disease had a higher fear level than those without one. This finding was consistent with that of a matched case-control study, which found that the prevalence of anxiety symptoms and depressive symptoms and the level of stress were significantly higher among those with preexisting chronic health conditions (59%, 71.6%, and 73.7%, respectively) compared with controls (25.6%, 31.1%, and 43.3%, respectively) [39]. Evidence has suggested that the presence of comorbid chronic conditions would increase the risk of death from COVID-19 [40-42]. Moreover, one major concern with the COVID-19 pandemic was its impacts on the routine use of health care services especially for individuals with comorbidities [43]. Service disruptions due to cancellations of elective care and lockdowns hindering access to health care facilities, in addition to the diffidence of patients with a chronic disease in seeking assistance for fear of risking iatrogenic exposure, altogether increased the psychological burden of patients with

a chronic disease. Thus, it was not surprising that people with a chronic disease had a higher fear level than those without.

In this study, more than 40% of the study participants reported that the thought of COVID-19 scared them. About one-third of the study participants reported that when they thought about COVID-19, they felt nervous, uneasy, and depressed. No doubt, the COVID-19 pandemic was very stressful for people and the communities in general [7]; the fear of infection was very common during the pandemic. Furthermore, people were worried that the health care system could not cope with the COVID-19 pandemic, that there were not enough hospital beds and ventilators to handle the rising number of COVID-19 cases. Another concern weighing on people's minds was the COVID-19 recession. Fear of the COVID-19 pandemic can be overwhelming and cause strong emotions [7]. It was also noteworthy that the COVID-19 pandemic rekindled fears of the 2003 severe acute respiratory syndrome epidemic in mainland China and Hong Kong.

Implications

First, based on the psychometric evaluation, we found that the adapted scale was a valid and reliable measure to assess the level of fear related to COVID-19. Further studies can use this scale to longitudinally monitor the fear level in different communities. Second, given the high fear levels found in the study sample, it is required to provide psychosocial care for the general public to diminish the psychological burden of the pandemic. Third, the findings call for the need to provide more psychosocial care for chronic disease patients and older adults.

Conclusion

This study found that the psychometric properties of the Fear Scale were acceptable to evaluate the fear level of the general Chinese population. Our descriptive analysis found that more than 40% of the study participants reported nervousness when they thought about COVID-19. About one-third of the study participants reported that when they thought about COVID-19, they felt nervous, uneasy, and depressed. Additionally, we found that people with a chronic disease reported a higher fear level than those without. The findings call for the need to provide more psychosocial care for chronic disease patients and older adults.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 30 KB - formative_v6i3e31992_app1.docx]

References

1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
2. Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. Acta Biomed 2020 Mar 19;91(1):157-160 [FREE Full text] [doi: 10.23750/abm.v91i1.9397] [Medline: 32191675]
3. COVID-19 Dashboard. John Hopkins University. URL: https://coronavirus.jhu.edu/map.html [accessed 2022-02-18]
4. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2022-02-18]
5. Lam HY, Lau CCA, Wong CH, Lee KYK, Yip SL, Tsang KLA, et al. A review of epidemiology and public health control measures of COVID-19 variants in Hong Kong, December 2020 to June 2021. IJID Regions 2022 Mar;2:16-24. [doi: 10.1016/j.ijregi.2021.11.002]
6. Qiu J, Shen B, Zhao M, Wang Z, Xie B, Xu Y. A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations. Gen Psychiatr 2020;33(2):e100213 [FREE Full text] [doi: 10.1136/gpsych-2020-100213] [Medline: 32215365]
7. Choi EPH, Hui BPH, Wan EYF. Depression and Anxiety in Hong Kong during COVID-19. Int J Environ Res Public Health 2020 May 25;17(10):3740 [FREE Full text] [doi: 10.3390/ijerph17103740] [Medline: 32466251]

8.   Choi EPH, Hui BPH, Wan EYF, Kwok JYY, Tam THL, Wu C. COVID-19 and Health-Related Quality of Life: A Community-Based Online Survey in Hong Kong. Int J Environ Res Public Health 2021 Mar 20;18(6):3228 [FREE Full text] [doi: 10.3390/ijerph18063228] [Medline: 33804725]

9.   Allen R, Prescott J, McHugh S, Carson J. Loneliness and mental health at the early stages of the Covid-19 pandemic in England. Health Soc Care Community 2021 Dec 09:1-11. [doi: 10.1111/hsc.13676] [Medline: 34888972]

10.   Maleku A, Kim YK, Kirsch J, Um MY, Haran H, Yu M, et al. The hidden minority: Discrimination and mental health among international students in the US during the COVID-19 pandemic. Health Soc Care Community 2021 Dec 17:Online ahead of print. [doi: 10.1111/hsc.13683] [Medline: 34921449]

11.   Cénat JM, Blais-Rochette C, Kokou-Kpolou CK, Noorishad P, Mukunzi JN, McIntee S, et al. Prevalence of symptoms of depression, anxiety, insomnia, posttraumatic stress disorder, and psychological distress among populations affected by the COVID-19 pandemic: A systematic review and meta-analysis. Psychiatry Res 2021 Jan;295:113599 [FREE Full text] [doi: 10.1016/j.psychres.2020.113599] [Medline: 33285346]

12.   Salari N, Hosseinian-Far A, Jalali R, Vaisi-Raygani A, Rasoulpoor S, Mohammadi M, et al. Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. Global Health 2020 Jul 06;16(1):57 [FREE Full text] [doi: 10.1186/s12992-020-00589-w] [Medline: 32631403]

13.   Ahorsu DK, Lin C, Imani V, Saffari M, Griffiths MD, Pakpour AH. The Fear of COVID-19 Scale: Development and Initial Validation. Int J Ment Health Addict 2020 Mar 27:1-9 [FREE Full text] [doi: 10.1007/s11469-020-00270-8] [Medline: 32226353]

14.   Dunsmoor JE, Paz R. Fear Generalization and Anxiety: Behavioral and Neural Mechanisms. Biol Psychiatry 2015 Sep 01;78(5):336-343. [doi: 10.1016/j.biopsych.2015.04.010] [Medline: 25981173]

15.   Han MFY, Mahendran R, Yu J. Associations Between Fear of COVID-19, Affective Symptoms and Risk Perception Among Community-Dwelling Older Adults During a COVID-19 Lockdown. Front Psychol 2021 Mar 23;12:638831 [FREE Full text] [doi: 10.3389/fpsyg.2021.638831] [Medline: 33833722]

16.   Harper C, Satchell L, Fido D, Latzman R. Functional Fear Predicts Public Health Compliance in the COVID-19 Pandemic. Int J Ment Health Addict 2021;19(5):1875-1888 [FREE Full text] [doi: 10.1007/s11469-020-00281-5] [Medline: 32346359]

17.   Ho CS, Chee CY, Ho RC. Mental Health Strategies to Combat the Psychological Impact of Coronavirus Disease 2019 (COVID-19) Beyond Paranoia and Panic. Ann Acad Med Singap 2020 Mar 16;49(3):155-160 [FREE Full text] [Medline: 32200399]

18.   Dsouza DD, Quadros S, Hyderabadwala ZJ, Mamun MA. Aggregated COVID-19 suicide incidences in India: Fear of COVID-19 infection is the prominent causative factor. Psychiatry Res 2020 Aug;290:113145 [FREE Full text] [doi: 10.1016/j.psychres.2020.113145] [Medline: 32544650]

19.   Arafat SY, Kar SK, Marthoenis M, Sharma P, Hoque Apu E, Kabir R. Psychological underpinning of panic buying during pandemic (COVID-19). Psychiatry Res 2020 May 06;289:113061. [doi: 10.1016/j.psychres.2020.113061] [Medline: 33242817]

20.   Sit S, Lam T, Lai A, Wong B, Wang M, Ho S. Fear of COVID-19 and its associations with perceived personal and family benefits and harms in Hong Kong. Transl Behav Med 2021 Apr 07;11(3):793-801 [FREE Full text] [doi: 10.1093/tbm/ibab018] [Medline: 33755146]

21.   Champion VL, Skinner CS, Menon U, Rawl S, Giesler RB, Monahan P, et al. A breast cancer fear scale: psychometric development. J Health Psychol 2004 Nov 01;9(6):753-762. [doi: 10.1177/1359105304045383] [Medline: 15367754]

22.   Leung DY, Wong EM, Chan CW. Adapting Champion's Breast Cancer Fear Scale to colorectal cancer: psychometric testing in a sample of older Chinese adults. Eur J Oncol Nurs 2014 Jun;18(3):281-285. [doi: 10.1016/j.ejon.2014.01.007] [Medline: 24629503]

23.   Ferguson KN, Coen SE, Tobin D, Martin G, Seabrook JA, Gilliland JA. The mental well-being and coping strategies of Canadian adolescents during the COVID-19 pandemic: a qualitative, cross-sectional study. CMAJ Open 2021 Nov 16;9(4):E1013-E1020 [FREE Full text] [doi: 10.9778/cmajo.20210042] [Medline: 34785531]

24.   Šrol J, Ballová Mikušková E, Čavojová V. When we are worried, what are we thinking? Anxiety, lack of control, and conspiracy beliefs amidst the COVID-19 pandemic. Appl Cogn Psychol 2021 Feb 09;35(3):720-729 [FREE Full text] [doi: 10.1002/acp.3798] [Medline: 33821088]

25.   Ballová Mikušková E, Verešová M. Distance education during COVID-19: The perspective of Slovak teachers. PEC 2020 Dec 5;78(6):884-906. [doi: 10.33225/pec/20.78.884]

26.   Lok KY, Fong DYT, Wong JY, Ho M, Choi EP, Pandian V, CARE group. International survey for assessing COVID-19's impact on fear and health: study protocol. BMJ Open 2021 May 11;11(5):e048720 [FREE Full text] [doi: 10.1136/bmjopen-2021-048720] [Medline: 33980534]

27.   Deek H, El Nayal M, Alabdulwahhab K, Ahmad M, Shaik R, Alzahrani M, et al. A multi-centric study on validation of the Fear Scale for COVID-19 in five Arabic speaking countries. Brain Behav 2021 Nov 17;11(11):e2375 [FREE Full text] [doi: 10.1002/brb3.2375] [Medline: 34661971]

28.   CARE: An international survey for assessing COVID-19's impact on fear and health. School of Nursing, The University of Hong Kong. URL: https://www.care.hku.hk [accessed 2022-02-17]

29. Guo N, Luk TT, Ho SY, Lee JJ, Shen C, Oliffe J, et al. Problematic Smartphone Use and Mental Health in Chinese Adults: A Population-Based Study. Int J Environ Res Public Health 2020 Jan 29;17(3):844 [FREE Full text] [doi: 10.3390/ijerph17030844] [Medline: 32013111]

30. Dunlow N, Phillips C, Broder HL. Concurrent validity of the COHIP. Community Dent Oral Epidemiol 2007 Aug;35 Suppl 1(s1):41-49 [FREE Full text] [doi: 10.1111/j.1600-0528.2007.00404.x] [Medline: 17615049]

31. Feng S, Zhang Q, Ho SMY. Fear and anxiety about COVID-19 among local and overseas Chinese university students. Health Soc Care Community 2021 Nov 06;29(6):e249-e258 [FREE Full text] [doi: 10.1111/hsc.13347] [Medline: 33825220]

32. Sánchez-Teruel D, Robles-Bello MA, Lara-Cabrera M, Valencia-Naranjo N. Gender implications of the Fear of COVID-19 Scale in the Spanish population: A validation study. Psychol Trauma 2022 Feb;14(2):258-265. [doi: 10.1037/tra0001062] [Medline: 34180687]

33. Fan HSL, Choi EPH, Ko RWT, Kwok JYY, Wong JYH, Fong DYT, et al. COVID-19 related fear and depression of pregnant women and new mothers. Public Health Nurs 2021 Dec 13:Online ahead of print. [doi: 10.1111/phn.13035] [Medline: 34902184]

34. Özmen S, Özkan O, Özer Ö, Yanardağ MZ. Investigation of COVID-19 Fear, Well-Being and Life Satisfaction in Turkish Society. Soc Work Public Health 2021 Feb 17;36(2):164-177. [doi: 10.1080/19371918.2021.1877589] [Medline: 33461428]

35. Witte K. Putting the fear back into fear appeals: The extended parallel process model. Communication Monographs 1992 Dec;59(4):329-349. [doi: 10.1080/03637759209376276]

36. Rosenhan DL, Seligman MEP. Abnormal psychology. New York, US: WW Norton & Co; 1995.

37. Gullestad SE. One depression or many? The Scandinavian Psychoanalytic Review 2013 Jan 21;26(2):123-130. [doi: 10.1080/01062301.2003.10592920]

38. Sylvers P, Lilienfeld SO, LaPrairie JL. Differences between trait fear and trait anxiety: implications for psychopathology. Clin Psychol Rev 2011 Feb;31(1):122-137. [doi: 10.1016/j.cpr.2010.08.004] [Medline: 20817337]

39. Sayeed A, Kundu S, Al Banna MH, Christopher E, Hasan M, Begum M, et al. Mental Health Outcomes of Adults with Comorbidity and Chronic Diseases during the COVID-19 Pandemic: A Matched Case-Control Study. Psychiatr Danub 2020;32(3-4):491-498 [FREE Full text] [doi: 10.24869/psyd.2020.491] [Medline: 33370758]

40. Targher G, Mantovani A, Wang X, Yan H, Sun Q, Pan K, et al. Patients with diabetes are at higher risk for severe illness from COVID-19. Diabetes Metab 2020 Sep;46(4):335-337 [FREE Full text] [doi: 10.1016/j.diabet.2020.05.001] [Medline: 32416321]

41. Onder G, Rezza G, Brusaferro S. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. JAMA 2020 May 12;323(18):1775-1776. [doi: 10.1001/jama.2020.4683] [Medline: 32203977]

42. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. Int J Infect Dis 2020 May;94:91-95 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.017] [Medline: 32173574]

43. Wang J, Spencer A, Hulme C, Corbett A, Khan Z, Vasconcelos Da Silva M, et al. Healthcare utilisation and physical activities for older adults with comorbidities in the UK during COVID-19. Health Soc Care Community 2021 Dec 09:Online ahead of print. [doi: 10.1111/hsc.13675] [Medline: 34888982]

Abbreviations

PHQ: Patient Health Questionnaire

XSL•FO

RenderX

<u>Original Paper</u>

Social Media for ImpLementing Evidence (SMILE): Conceptual Framework

Junqiang Zhao[1], MScN; Gillian Harvey[2], PhD; Amanda Vandyk[1], PhD; Wendy Gifford[1], PhD

[1]School of Nursing, Faculty of Health Sciences, University of Ottawa, Ottawa, ON, Canada
[2]College of Nursing and Health Sciences, Flinders University, Adelaide, Australia

Corresponding Author:
Junqiang Zhao, MScN
School of Nursing
Faculty of Health Sciences
University of Ottawa
451 Smyth Rd
Ottawa, ON, K1H 8M5
Canada
Phone: 1 8192124678
Email: jzhao126@uottawa.ca

Abstract

Background: Social media has become widely used by individual researchers and professional organizations to translate research evidence into health care practice. Despite its increasing popularity, few social media initiatives consider the theoretical perspectives of how social media works as a knowledge translation strategy to affect research use.

Objective: The purpose of this paper is to propose a conceptual framework to understand how social media works as a knowledge translation strategy for health care providers, policy makers, and patients to inform their health care decision-making.

Methods: We developed this framework using an integrative approach that first involved reviewing 5 long-standing social media initiatives. We then drafted the initial framework using a deductive approach by referring to 5 theories on social media studies and knowledge translation. A total of 58 empirical studies on factors that influenced the use of social media and its messages and strategies for promoting the use of research evidence via social media were further integrated to substantiate and fine-tune our initial framework. Through an iterative process, we developed the Social Media for ImpLementing Evidence (*SMILE*) framework.

Results: The *SMILE* framework has six key constructs: developers, messages and delivery strategies, recipients, context, triggers, and outcomes. For social media to effectively enable recipients to use research evidence in their decision-making, the framework proposes that social media content developers respond to target recipients' needs and context and develop relevant messages and appropriate delivery strategies. The recipients' use of social media messages is influenced by the virtual–technical, individual, organizational, and system contexts and can be activated by three types of triggers: sparks, facilitators, and signals.

Conclusions: The *SMILE* framework maps the factors that are hypothesized to influence the use of social media messages by recipients and offers a heuristic device for social media content developers to create interventions for promoting the use of evidence in health care decision-making. Empirical studies are now needed to test the propositions of this framework.

Introduction

Social Media Use in Health Care

Social media has been extensively used worldwide to communicate health-related information. For example, in China, one-third of the users of the social media platform WeChat—which is widely used for instant messaging and social networking [1]—receive and read health information through the platform [2]. In the United States, 32% of social media users post messages about friends and family members' health experiences on social media [3]. Health care professionals use social media to provide health information and answer medical

questions [4], and patients and caregivers use social media for self-care and health literacy [5]. In health care research, social media platforms such as Twitter, Facebook, and YouTube are increasingly used for participant recruitment, intervention implementation, data mining and collection, and the sharing of research findings [6].

Social media, with its free access, interactive features, and widespread reach, has become increasingly used by individual researchers and professional organizations who wish to translate research evidence into health care practice. For example, the Joanna Briggs Institute (JBI) at Fudan University in China has been using WeChat to disseminate nursing evidence since 2014. In the first 2 years, their WeChat account reached 22,369 followers from 34 provinces in China [7,8]. The Cochrane Child Health groups in Canada and Portugal used social media strategies to disseminate child health evidence to health care providers, and within 6 months of initiating the strategy, their blog received 2555 visitors and 3967 page views, and their Twitter account gained 469 new followers from a geographically diverse population [9]. A social media initiative called *It Doesn't Have to Hurt*, led by health care researchers in Canada, developed a short YouTube video on evidence-based strategies, such as distraction and using topical anesthetics for reducing procedural pain in children. Their video received 237,132 unique views from 182 countries 5 years after its launch, with patients and health care providers reporting strong acceptance and high intention to use the strategies [10]. The number of parents reporting the use of topical anesthetic creams to reduce pain increased from 18% to 63% after watching the video [11].

There has also been a surge in social media initiatives during the COVID-19 pandemic, which are aimed at helping health care professionals, patients, and the public better understand the coronavirus and cope with its impacts. Global evidence synthesis networks such as Cochrane, JBI, and Campbell Collaboration use social media to disseminate rapid review findings related to COVID-19. In China, the Beijing University of Chinese Medicine (BUCM) Cochrane Center, together with 20 evidence-based health care research teams and organizations, launched the *Fighting COVID-19 with Evidence* initiative. They collect urgent clinical questions about COVID-19 diagnosis, treatment, and nursing care through WeChat and share recommendations after a rapid search and synthesis of research evidence [12]. In England, the Center for Evidence-Based Medicine at Oxford University uses Twitter (@CebmOxford) to share COVID-19 relevant recommendations to a global audience. In Canada, the COVID-19 Evidence Network to support Decision-making initiative (@COVID_E_N_D) collects the best available evidence related to COVID-19 and shares this information on Twitter to support decision-making.

Theoretical Understandings of Social Media as a Knowledge Translation Strategy

Despite its popularity, many researchers and organizational decision-makers upload research findings onto social media platforms without deliberately planning how to facilitate its use by recipients in policies, programs, or practices. In their systematic review, Webb et al [13] concluded that theory-based internet interventions had greater impacts on health behaviors

than non–theory-based interventions, with interventions based on the *theory of planned behavior* having larger effects than those based on the *transtheoretical model* or *social cognitive theory*. However, despite these benefits, theoretical frameworks are rarely used to guide the development of social media interventions aimed at facilitating research use. In their systematic review, Arguel et al [14] only identified 15 experimental studies published between 2005 and 2016 that applied theoretical approaches to guide the development of social media interventions.

Ngai et al [15,16] classified 31 theories used in social media studies into three categories: personal behavior theories, social behavior theories, and mass communication theories. Personal behavior theories (eg, the *theory of planned behavior* and *technology acceptance model*) focus on personal factors that affect user behavior on social media. Social behavior theories (eg, *social capital theory* and *social cognitive theory*) identify key social factors that stimulate individuals to participate in collective actions on social media. Mass communication theories (eg, *parasocial interaction theory*) reveal the distinct characteristics of social communications that can assist in the use of social media for communication and marketing [15,16]. These theories provide valuable insights into social media's role in behavior change; however, the following two limitations exist in fully understanding the research use process:

1. They only consider 1 of the 2 latent and indispensable layers of social media use: social media and messages. Recipients must first use social media before they can engage with messages (eg, the *technology acceptance model* emphasizes the platform, and the *social cognitive theory* and *theory of planned behavior* focus on the message). Theories that do not address both layers fail to fully explain the process of research use through social media.
2. They neglect multilevel contextual factors, such as the virtual–technical, organizational, and system contexts, particularly in relation to the features of the social media platform in shaping behavior. This may lead to the development of knowledge translation strategies solely from an individual perspective, without taking into account the contextual determinants that affect recipients' behaviors.

These 2 limitations were partially addressed by Ritterband et al [17], who developed a *behavior change model for internet interventions*, which posited that website use was influenced by support, characteristics of the websites and users, and environmental factors. Behavior change from information on websites is then influenced by various mechanisms (eg, knowledge and motivation). This model has been used to guide the development and evaluation of internet interventions in health care [18,19]. Although not exactly the same, websites that allow for multiway interaction are normally considered to be social media [20,21], and the Ritterband et al [17] model has been used in the social media context [22]. It addresses the limitations of the aforementioned social media theories, as it considers the platform—which in this case is the website—and accounts for the multilayered contexts in shaping behavior, such as personal, professional, and community contexts, as well as the health care system [17]. However, the Ritterband et al model [17] does not make mechanisms of change explicit and presents

a linear process for using the internet to change behavior when real-world practice is often complex [17].

Despite its extensive use for disseminating health care research evidence, social media is rarely used in a well-planned way with end users in mind, which largely limits its potential to bridge evidence–practice gaps and contribute to health care practices. Studies on the use of research evidence through social media are sparse [14]. Large theoretical gaps exist in understanding how social media interventions affect health care practices and decision-making. Unpacking the process by which social media works as a knowledge translation strategy is important to not only advance science but also inform interventions for improving health care practices and patient outcomes.

Objective

The purpose of this paper is to propose a conceptual framework to understand how social media works as a knowledge translation strategy for health care providers, policy makers, and patients to inform their health care decision-making.

Methods

We used a 3-step process based on the approach described by Meleis [23] to develop our conceptual framework. Meleis suggested that practice, theory, and research are important sources for patterning real-world phenomena and informing theory development [23]. Our approach was iterative and flexible and built a preliminary understanding of the process through which social media works for knowledge translation.

To get a sense of how they operate, we first reviewed five long-standing social media initiatives that have a large number of followers: the Fudan University JBI Center Nursing Evidence Dissemination Initiative (ie, Fudan JBI Initiative) [7,8,24], BUCM Cochrane Evidence Dissemination Initiative (ie, BUCM Cochrane Initiative) [25], *It Doesn't Have to Hurt* initiative [10,26], *Be Sweet to Babies* initiative [27], and Translating Evidence in Child Health to Enhance Outcomes (ECHO) program [28]. For each social media initiative, we specifically reviewed the topics and interface of their social media channels (including format and structure of content); the number of readers, followers, and comments; intervals between posts; and the length of videos and papers published relating to each initiative.

Second, we drafted the initial framework using a deductive approach based on existing theories and our team members' expertise in knowledge translation and social media. We primarily drew on five well-known and widely cited theories, frameworks, and models: *integrated Promoting Action on Research Implementation in Health Services (i-PARIHS)* [29]; *capability, opportunity, motivation, and behavior (COM-B)* [30]; *Fogg behavior model* [31]; *theory of innovation diffusion* [32]; and *behavior change model for Internet interventions* [17]. We built the basic structure of our framework based on the *i-PARIHS framework*, which argues that successful knowledge translation relies on the interactions among four constructs: innovation, recipients, context, and facilitation. In addition to the *i-PARIHS* constructs, we added one construct for social media content developers (hereafter referred to as *developers*)

in recognition of the crucial role they play in ensuring that recipients get relevant and appropriate messages. We added the virtual–technical context to the 3-layer contexts described in *i-PARIHS* (ie, local, organizational, and external) to capture the unique features of social media platforms, which is substantiated by the *behavior change model for internet interventions* (described above). We also included three types of knowledge translation outcomes—conceptual, instrumental, and persuasive research use [33,34]—in recognition of the fact that not all evidence on social media was appropriate for practice or behavior change. Rather, we recognize that a large amount of social media evidence affects understanding, attitudes, or collective actions.

The other four theories and models were used to develop two further aspects of our framework: using social media and using the messages. In the first aspect, the four theories and models were employed to understand social media use from two main construct levels: recipients and the virtual–technical context. In the second aspect, derived from the *COM-B model* and the *Fogg behavior model*, we built subconstructs for the active ingredient of message use, named as *trigger* in our framework.

We then reviewed published papers that incorporated the 5 long-standing social media initiatives (described earlier) and used strategies such as citation tracking from the papers we reviewed. The forward citation search was conducted using Google Scholar, and the backward citation search was conducted by screening the reference lists. We also conducted a citation snowballing search using Google Scholar and consulted experts from the 5 social media initiatives and our team members to further locate relevant empirical studies. The studies we identified were primarily about factors that influenced people's use of social media and its messages and strategies for promoting message use. We used the key findings of these studies to substantiate and fine-tune our initial framework. Through an iterative process, we went back and forth from the initial framework to social media initiatives, theories, and empirical studies and developed the Social Media for ImpLementing Evidence (*SMILE*) framework. *Implementation* in the *SMILE* framework refers to instrumental, conceptual, and persuasive knowledge translation.

Results

Overview

Through a review of social media initiatives (n=5), theories (n=5), and empirical studies (n=58), including papers (15/58, 26%) relevant to the 5 social media initiatives [7,8,10,24,26,27,35-43] and papers (43/58, 74%) [9,13,15,44-83] from citation tracking, snowballing, or consultation, we developed the *SMILE* framework (Figure 1). Table 1 summarizes the key constructs and their supporting evidence. The *SMILE* framework provides a preliminary understanding of how social media can be used as a knowledge translation strategy to inform health care practices and decision-making. It has six key constructs: (1) developers, (2) messages and delivery strategies, (3) recipients, (4) context, (5) triggers, and (6) outcomes. For social media to enable recipients to use research evidence in their practice or decision-making, the framework proposes that

developers respond to the needs and context of target recipients to develop relevant messages and appropriate delivery strategies. Recipients' use of social media messages is influenced by the virtual–technical, individual, organizational, and system contexts and can be activated by different types of triggers, described as sparks, facilitators, and signals. Next, we describe the constructs of the *SMILE* framework.

Figure 1. SMILE (Social Media for ImpLementing Evidence) framework.
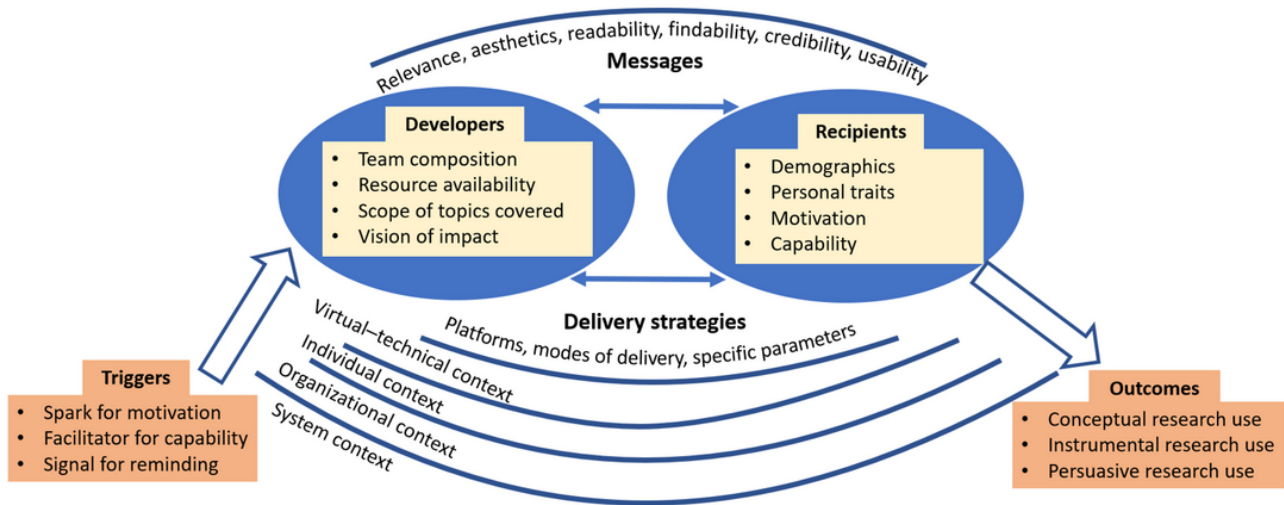
Table 1. Key constructs in the Social Media for ImpLementing Evidence (SMILE) framework and the supporting evidence.

Constructs	Theory origins	Empirical studies	Social media initiatives
Developers	—[a]	[9,78]	Fudan JBI[b] Initiative [7,8,24,84]; BUCM[c] Cochrane Initiative [25]; *It Doesn't Have to Hurt* initiative [10,26,85]; *Be Sweet to Babies* initiative [27]; ECHO[d] [28,43]
Team composition		[9]	Fudan JBI Initiative [7,8,24]; BUCM Cochrane Initiative [25]; *It Doesn't Have to Hurt* initiative [85]; *Be Sweet to Babies* initiative [27]; ECHO [28]
Resource availability		[9]	Fudan JBI Initiative [7,8,24]; BUCM Cochrane Initiative [25]; *It Doesn't Have to Hurt* initiative [10,26]; *Be Sweet to Babies* initiative [27]; ECHO [28]
Scope of topic covered		[9]	Fudan JBI Initiative [7,8,24]; BUCM Cochrane Initiative [25]; *It Doesn't Have to Hurt* initiative [10,26]; *Be Sweet to Babies* initiative [27]; ECHO [28,43]
Vision of impact		[51,79]	Fudan JBI Initiative [84]; BUCM Cochrane Initiative [25]; *It Doesn't Have to Hurt* initiative [10,26,85]; *Be Sweet to Babies* initiative [27]; ECHO [28,43]
Messages and delivery strategies	*i-PARIHS[e] framework* (innovation) [29]	[51,57]	Fudan JBI Initiative [7,8,24]; BUCM Cochrane Initiative [25]; *It Doesn't Have to Hurt* initiative [10,26]; *Be Sweet to Babies* initiative [27]; ECHO [28,43]
Messages	*Behavior change model for internet interventions* (website) [17]	[44,47,49,52,54,56,58-63,80]	Fudan JBI Initiative [7,8,24]; BUCM Cochrane Initiative [25]; *It Doesn't Have to Hurt* initiative [10,26]; *Be Sweet to Babies* initiative [27]; ECHO [28,43]
Delivery strategies	*Behavior change model for internet interventions* (website) [17]	[13,82,83]	Fudan JBI Initiative [7,8,24]; BUCM Cochrane Initiative [25]; *It Doesn't Have to Hurt* initiative [10,26]; *Be Sweet to Babies* initiative [27]; ECHO [28,43]
Recipients	*i-PARIHS framework* (recipients) [29]	[15,48,55,65]	—
Demographics	*Behavior change model for internet interventions* (user characteristics [17]	[15,45,53,55]	
Personal traits	*Behavior change model for internet interventions* (user characteristics) [17]	[46,64]	
Motivation	*COM-B[f] model* (motivation and capability) [30]; *Fogg behavioral model* (motivation and capability) [31]	[48,50,55,65]	
Capability	*COM-B model* (motivation and capability) [30]; *Fogg behavioral model* (motivation and capability) [31]	[48,50,55,65]	
Context	*i-PARIHS framework* (context) [29]	[81]	
Virtual–technical context	*Behavior change model for internet interventions* (website) [17]; *theory of innovation diffusion* (innovation characteristics) [32]	[48,65,69,70,75,77]	Fudan JBI Initiative [7,8,24]; BUCM Cochrane Initiative [25]; *It Doesn't Have to Hurt* initiative [10,26]; *Be Sweet to Babies* initiative [27,39,40]; ECHO [28]

Constructs	Theory origins	Empirical studies	Social media initiatives
Individual context	*Behavior change model for internet interventions* (environment) [17]; *COM-B model* (environment) [30]	[67,68]	Fudan JBI Initiative [8,24]; *It Doesn't Have to Hurt* initiative [10,26]; *Be Sweet to Babies* initiative [39,40]
Organizational context	*Behavior change model for internet interventions* (environment) [17]; *COM-B model* (environment) [30]	—	*It Doesn't Have to Hurt* initiative [10,26]; *Be Sweet to Babies* initiative [39,40]
System context	*Behavior change model for internet interventions* (environment) [17]; *COM-B model* (environment) [30]	[71]	*Be Sweet to Babies* initiative [39,40]
Triggers	*i-PARIHS framework* (facilitation) [29]	[13,66,73,74]	—
Spark for motivation	*Fogg behavioral model* (trigger) [31]; *COM-B model* (motivation and capability) [30]; *behavior change model for internet interventions* (support) [17]	[55,67,74,79]	*Be Sweet to Babies* initiatives [41]
Facilitator for capacity	*Fogg behavioral model* (trigger) [31]; *COM-B model* (motivation and capability) [30]; *behavior change model for internet interventions* (support) [17]	[13,72,76]	*Be Sweet to Babies* initiatives [27,42]
Signal for reminding	*Fogg behavioral model* (trigger) [31]; *COM-B model* (motivation and capability) [30]; *behavior change model for internet interventions* (support) [17]	[13,55,74]	—
Outcomes	*i-PARIHS framework* (successful implementation) [29]	[33,34]	—
Conceptual research use	—	—	*It Doesn't Have to Hurt* initiative [10,26]; *Be Sweet to Babies* initiative [27,35-39];
Instrumental research use	—	[76]	*Be Sweet to Babies* initiative [41]
Persuasive research use	—	—	—

[a]Data not available.

[b]JBI: Joanna Briggs Institute.

[c]BUCM: Beijing University of Chinese Medicine.

[d]ECHO: Evidence in Child Health to Enhance Outcomes.

[e]i-PARIHS: integrated Promoting Action on Research Implementation in Health Services.

[f]COM-B: capability, opportunity, motivation, and behavior.

Developers

Developers are individuals, groups, and organizations responsible for the management of social media contents. Developer activities may include designing and periodic uploading of information, monitoring operations, collecting data on impact, and answering questions or comments from viewers. Developers can be health care researchers who produce research evidence and share it directly via social media for public access. Barton [78] proposed a new research-to-practice continuum where researchers not only disseminate research findings through traditional journal publications but also create multimedia messages and disseminate them to the public. Developers can also be intermediaries who serve as a link between research producers and end users by translating research evidence into user-friendly messages for dissemination on social media.

Although it might be simple for individuals to develop and upload research findings to social media, a fast, frugal, and *hope-the-change-happens* approach has limitations. One of the propositions embedded in the framework is that the composition of the development team, availability of resources, scope of topics, and vision of impact influence the development of relevant and appropriate social media interventions, thus affecting recipients' engagement with and use of the messages.

We suggest bringing together a multidisciplinary collaborative team of health care professionals, target users, social media experts, and audiovisual technicians (eg, camera operators and video editors) to best support the development of social media

interventions [9]. Health care professionals can assist with the identification of different types of evidence resources; target users can strengthen the relevance and accessibility of messages; social media experts can contribute to the operation of the platform; and audiovisual technicians can provide support when the team wants to deliver messages using videos or animations. For example, the *It Doesn't Have to Hurt* initiative has built a large interdisciplinary collaborative team of researchers, trainees, patients, and other stakeholders to facilitate the stable operation of their social media program [10,26,85]. Similarly, the ECHO research program has created various videos, animations, and posters on child health with a multidisciplinary team [28]. With different knowledge, skills, and perspectives, the team can generate high-quality and influential social media products. The long-term collaborative approach can additionally promote the sustainability of these initiatives.

The availability of resources to develop and manage social media initiatives, such as time and budget, must be taken into consideration when planning it. In the *It Doesn't Have to Hurt* initiative, it cost the team Can $15,000 (US $11,802) and considerable efforts to develop and promote their YouTube video, and the developers stated that financial and time costs could be a hindrance for individual researchers to undertake the work [10]. In their social media initiative to disseminate Cochrane Child Health evidence, Dyson et al [9] also found that the team invested enormous time and human resources in managing the platform. Therefore, we suggest that adequate time and budget be allocated to social media initiatives before their commencement.

The scope of topics covered is closely linked to the amount of time and resources invested. Some initiatives, such as the Fudan JBI Initiative [7,8,24] and BUCM Cochrane Initiative [25], have broad scopes that are open to a range of topics in nursing and medicine. Some initiatives focus only on specific topics; for example, the *It Doesn't Have to Hurt* [10,26] and *Be Sweet to Babies* initiatives [27] target reducing procedural pain for children and infants, respectively. Other initiatives center on a certain field, such as the ECHO initiative, which covers common childhood conditions. The topics covered should be balanced with the consideration of practical issues. Dyson et al [9] suggested that starting from a specific content area and engaging with a stable social media community was more effective for developing a social media network.

It is also essential that the development team builds a shared vision of the impact they are looking to achieve and tracks the performance of their social media initiatives [51]. Building and sustaining a social media initiative is demanding work that requires collaboration and investment. An explicit team vision of the impact of social media can motivate the team to work toward a common goal. For example, since 2016, the Fudan JBI Initiative has openly shared its social media vision in its annual center report and at conferences [84]. Gates et al [79] also emphasized the importance of setting goals and tracking achievements after the evaluation of their social media initiative.

In the *SMILE* framework, we propose that the engagement of a multidisciplinary team, time, and resource investments are essential for developing relevant and appropriate social media

interventions to influence research use. Developers should balance the topics covered with practical considerations and create a shared vision of the goals of their social media initiatives.

Messages and Delivery Strategies

Overview

The second construct in the *SMILE* framework is messages and delivery strategies. Developers should respond to recipients' needs and their context to create messages and delivery strategies. Through a systematic literature review, Schein et al [57] observed that collaborating with target users to create social media interventions contributed to heightened authenticity of messages and improved trust in developers. Korda and Itani [51] suggested that social media messages should account for user characteristics and information preferences and should be customized through an iterative interaction with target users. On the basis of 4 years of experience in social media operations, the Fudan JBI Initiative recommended that developers could improve the usability and uptake of research evidence on WeChat through the full use of WeChat's interactive functions to capture users' needs [7,24].

Messages

To date, a limited number of studies have investigated the attributes of social media messages that influence its uptake, despite the development of tools and models to assess the quality of web-based information [44,47,49,54,59,61-63,80]. On the basis of the content of these tools and models, as well as the unique features of social media platforms, we posit six interrelated attributes that influence the uptake of social media messages: relevance, aesthetics, readability, findability, credibility, and usability.

A *relevant* message is directly related, connected, or pertinent to target users. The more relevant messages are to the target users, the higher their level of engagement and the likelihood of being used. In their systematic review, Schubart et al [58] concluded that internet interventions that addressed the primary concerns of patients with chronic health conditions were the most successful.

An *aesthetic* message is characterized by the artistic design and visual appeal of the social media content; for instance, the layout of content, color and size of words, and graphics [17]. A first impression is made after a brief glimpse of the format and structure of content, and a user will quickly decide whether to stay on it or leave [56]. For example, ECHO uses art-based approaches, such as animations and e-books, to disseminate child health evidence on social media [43]. As many social media platforms impose restrictions on the design and presentation of messages, flexibility with visual appeal is often limited. For example, Twitter only allows 140 characters and 4 pictures per tweet.

A *readable* message is easy to follow. The US National Institutes of Health recommend that the readability of content on websites be at the sixth- to eighth-grade level [44]. Readability also encompasses accessibility and understandability. Health information that is hard to read will

be hard to understand and therefore remain inaccessible, particularly for people with low health literacy [52]. For example, the *It Doesn't Have to Hurt* initiative developed YouTube video storyboards and scripts in collaboration with a communication company, which was further verified by parents for its readability [10,26].

The messages must also be *findable*, meaning that they are easy to locate. Search boxes, navigation menus, and links are likely to improve the findability of health information on social media [62,63]. Both the Fudan JBI Initiative [7] and BUCM Cochrane Initiative used the navigation function in WeChat to organize and categorize the evidence sources, which allowed users to easily locate the specific evidence item they wanted.

A *credible* message refers to the trustworthiness of the message and is described as accurate, believable, and factual [59,63,80]. The *Journal of the American Medical Association* considers four elements to judge the credibility of medical information on the internet: currency of information; declaration of authorship; presentation of a list of references; and the disclosure of any conflicts of interest, funding, or sponsorship [60].

Finally, the *usability* of a message is the extent to which it can be actionable in practice. For the purpose of affecting research use, clear behavioral recommendations or prescriptions within the message can promote its usability [44]. Together, the six attributes of relevance, aesthetics, readability, findability, credibility, and usability influence the use of a social media message in practice.

Delivery Strategies

Delivery strategies are the ways through which social media messages are conveyed to recipients. We conceptualize them as comprising three distinct layers: the social media platforms, modes of delivery, and specific parameters. One of the first decisions that developers need to make is which social media platform to use. Although social media platforms have burgeoned in recent years, only a few are popular for disseminating health care information, such as Facebook, YouTube, and Twitter in Western countries and WeChat and Weibo in Asia. Messages are delivered on social media platforms through different modes of delivery, such as text, infographics, videos, audios, animations, vignettes, testimonials, and stories [17]. The modes of delivery differ in their impact on users' engagement with the messages, and research has found that visual abstracts attract a significantly greater number of engagements than basic texts [82,83]. Webb et al [13] conducted a systematic review in which they classified the modes of delivery of internet-based behavior change interventions into three types: automated functions (eg, automated tailored feedback), communicative functions (eg, access to an adviser to request advice), and the use of supplementary modes (eg, SMS text message). It should be noted that the options for the mode of delivery vary for different social media platforms. The specific parameters of the delivery strategy are the characteristics of the mode of delivery, such as the length of videos, size, color and limits of words, frequency, and interval of message uploading. In the 5 initiatives we reviewed, all used a variety of social media platforms such as WeChat, YouTube, and Twitter. In addition, they used diversified modes of delivery,

such as videos, podcasts, animations, stories, and texts, to deliver their social media messages.

Overall, the attributes of messages and delivery strategies affect the reach and successful use of messages by people and are a key construct in the *SMILE* framework. The 6 attributes of messages and the 3 layers of delivery strategies should be considered during the social media content development process to promote the likelihood of message use.

Recipients

Recipients are the target audience of social media messages and have the potential to direct, influence, or be affected by messages. In our framework, we consider health care providers, policy makers, and health care consumers as recipients. We also propose that using social media messages in health care decision-making involves two distinct, interconnected layers: using the social media and then using the message. It is a prerequisite for recipients to first accept and use the social media before they can engage with the messages. We distinguish between these 2 layers and consider the factors that influence each layer separately. We contend that the characteristics of recipients and the virtual–technical context are the two main domains that influence people's use of social media, and the individual, organizational, and system contextual domains shape the message use.

Together with frameworks from the social media and technology research field [15,48,55,65], the *i-PARIHS framework* [29], *behavior change model for internet interventions* [17], *COM-B model* [30], and *Fogg behavior model* [31] have provided valuable insights into the characteristics of recipients that influence social media use. On the Basis of their theoretical constructs, four aspects of recipients' characteristics were incorporated into our framework: demographics, personal traits, motivation, and capability.

Demographics include age, gender, geography, socioeconomic status, ethnicity, and lifestyles [15,17,55]. Large quantities of research data from Twitter and Facebook revealed differences in social media use by gender, ethnicity, and geography [45,53]. Personal traits of openness, conscientiousness, extraversion, agreeableness, and neuroticism—rooted in genetics—are perceived as one of the fundamental theories that explain personal behavior [15]. They are closely associated with social media use [64]. In a national survey in the United States, Correa et al [46] found that although extraversion and openness were positively related to social media use, emotional stability—a central measure of neuroticism—was a negative predictor. These findings differed by gender and age [46].

Motivation and capability are 2 summative characteristics of social media recipients that the *SMILE* framework identifies as affecting social media use. These characteristics are based on the *Fogg behavior model* [31] and the *COM-B model* [30]. Within motivation, perceived needs [65], attitude [50], intention [48,50,55], self-efficacy [17], and goals [55] are factors motivating individuals to use social media. Within capability [48,50], knowledge and skills [17,31] enable individuals to use social media. Together, all four characteristics of recipients (demographics, personal traits, motivation, and capability) are

determinants affecting social media use in the *SMILE* framework.

Context

Overview

In the *SMILE* framework, context is defined as "a set of characteristics and circumstances that consist of active and unique factors that surround the implementation... (It) interacts, influences, modifies and facilitates or constrains the intervention and its implementation" [81]. We identify four interrelated layers of contextual factors that influence social media use and further message use: virtual–technical, individual, organizational, and system contexts.

Virtual–Technical Context

The virtual–technical context is the context surrounding the social media platform. Dawot and Ibrahim [69] summarized its composition into three core elements: individual-level, conversation-level, and community-level elements. Through a systematic review, Elaheebocus et al [70] created a taxonomy of social media features that included identity representation, communication, peer grouping, data sharing, competition, activity data viewing, and web-based social networks.

We posit that seven characteristics of the platform influence social media use: relative advantage, complexity, observability, compatibility, usefulness, interactivity, and playfulness [48,65,75,77]. Relative advantage, complexity, observability, and compatibility originate from the *theory of innovation diffusion* [32] and are all considered important factors influencing social media use [65]. *Relative advantage* is the degree to which one social media platform is perceived to be better than other alternatives. *Complexity* is the extent to which social media is perceived as being difficult to use. *Observability* is the degree to which the benefits of social media use are visible to others. *Compatibility* is the degree to which social media is perceived as consistent with the existing values, past experiences, and needs of potential users [65]. Each of these factors is positively associated with social media use, except for complexity [75]; the more complex the social media is perceived, the lower the level of engagement by users. *Usefulness* is the degree to which social media can directly or indirectly benefit individual performance. Data show that usefulness can predict up to 62% of the intention to use social media [48]. *Interactivity* is the degree to which social media enables 2-way communication rather than 1-way transmission or distribution of information. Multiple research studies have demonstrated the positive effects of interactivity on social media use [75,77]. *Playfulness* is the hedonic value of social media and can influence the perceived usefulness and direct use of social media [48]. In the 5 social media initiatives included for developing the *SMILE* framework, all of them use popular platforms that contain these 7 characteristics, attesting to their importance. We posit that all 7 aspects of the platform in the virtual–technical context affect social media use.

Individual Context

The context of an individual plays a crucial role in shaping one's behavior of message use. Brouwer et al [67,68] found that being motivated to visit the web-based intervention, being curious about the content, and perceiving the web-based intervention as personally relevant were important influencers for participants to engage with the web-based intervention. In a qualitative and a cross-sectional study conducted by Hu et al [39,40] to understand the barriers of implementing the *Be Sweet to Babies* pediatric pain management strategies in China, they found that insufficient knowledge, beliefs, and self-efficacy of health care providers were common individual-level barriers hindering the implementation of social media messages in clinical practices by nurses.

Organizational Context

Organizational context is considered an indispensable layer of the context affecting one's use of a social media message in practice. In the *Be Sweet to Babies* initiative, the hierarchical managerial system, low authority of nurses, and staff shortage were factors impeding nurses from changing their practice and incorporating the evidence in China [40]. In the *It Doesn't Have to Hurt* initiative, researchers found that the cost for using topical anesthetic cream [10] and the unit routines of disallowing parental presence during painful procedures [26] hindered the implementation of pain management strategies for children. The Fudan JBI Initiative also stated explicitly in every WeChat post that users should consider the local context to determine the appropriateness of implementing the evidence.

System Context

People's use of social media messages in health care practices is also influenced by the broader system context, namely the social, political, economic, and cultural environment. From a social perspective, one study found that popular opinion leaders on the internet played a positive role in changing sexual behaviors among men who have sex with men [71]. Some countries impose restrictions at the judicial level on accessing certain social media, which may be attributed to ideological, political, or economic reasons. Culturally, Hu et al [40] found that the negatively escalating relationships between patients and health care professionals in China made nurses reluctant to introduce the *Be Sweet to Babies* pain management strategies, despite a strong evidence base for the practices.

As illustrated above, we have made distinctions between the four types of contexts that influence social media and its message use in the *SMILE* framework. Specifically, the virtual–technical context concerns the determinants of social media use; the individual, organizational, and system contexts are considered as the micro-, meso-, and macro-level factors shaping message use.

Triggers

Overview

The concept of the *trigger* in the *SMILE* framework describes the strategies adopted to activate social media message use. On the basis of the *i-PARIHS framework* [29], for social media to be effective in facilitating research use, there needs to be an active ingredient to energize the message implementation process, in addition to having relevant messages. The trigger is derived from the *Fogg behavioral model* [31] and includes

behavior change techniques (active triggers) or events (passive triggers) that activate a recipient to use social media messages. One behavior change technique, as an active trigger, is an "observable, replicable, and irreducible component of an intervention designed to alter or redirect causal processes that regulate behavior" [73]. Michie et al [73] created a behavior change technique taxonomy to standardize the reporting of the active content of behavior change interventions. These techniques have been widely adopted in social media interventions. Webb et al [13] found, in their systematic review, that internet interventions that incorporated more behavior change techniques had larger effects than interventions that incorporated fewer techniques. In a systematic review of the characteristics of internet-delivered healthy lifestyle promotion interventions, Brouwer et al [66] reported that feedback, interactive elements, and email or phone contact were the most commonly used techniques. In a recent systematic review in 2020, Simeon et al [74] conducted a detailed analysis of the behavior change techniques used in social media interventions. They found that 46 techniques had been used in the identified 71 studies. An event, as a passive trigger, is an emergent, unexpected, or accidental incident that pushes recipients to use social media messages in a passive way. These events require people to think and act in alternative ways, and social media provides relevant information to perform an alternative behavior. Fogg [31] classified triggers into three different types in persuasive technology design: sparks, facilitators, and signals. We adopted these 3 types of triggers and enriched their connotations in our framework, as discussed in the following sections.

Spark for Motivation

A spark is a trigger that motivates recipients to use a message. It can be used when recipients' motivation to use a social media message is low or needs to be further enhanced. Developers can apply various behavior change techniques such as problem solving, feedback and monitoring, and social support and reward [55,74] to help activate behavior change. For example, Modanloo [41] used motivational interviewing to improve parents' use of *Be Sweet to Babies* pain management strategies for infants during vaccination. Although no significant differences were found between the 2 comparative groups, approximately all participants used at least one strategy in the vaccination [41]. Through a Delphi study approach, Brouwer et al [67] identified that two behavior change techniques—the provision of tailored feedback on behavior and credible information source [73]—were related to an extended engagement with internet interventions. After implementing their social media initiative, Gates et al [79] suggested that web-based opinion leaders' endorsements would be a promising strategy for motivating recipients to use the messages.

Facilitator for Capability

A facilitator is a trigger that improves recipients' capability to use social media messages, such as knowledge and skills. Social media interventions that incorporate different behavior change techniques, such as instructions on behavior performance and demonstrating the behavior [73], are likely to improve the capability of recipients. In the Clinical Excellence Through

Social Media trial, Tunnecliff et al [76] linked every tendon management practice point on Twitter and Facebook to supplementary information to enhance the knowledge of recipients. Webb et al [13] found that the use of communicative functions within internet interventions to provide access to and schedule contacts with an adviser could have a small to medium effect on behavior. Developers made full use of the visualization function of a YouTube video in the *Be Sweet to Babies* initiative to demonstrate pain management techniques and help the recipients build skills [27]. Watching this video doubled the chance of using an analgesic strategy and increased breastfeeding 1.5 times and skin-to-skin care 4.6 times by parents in a nonrandomized pragmatic trial in Brazil [42].

Signal for Reminding

A signal indicates or reminds recipients of social media messages. This type of trigger is useful when recipients need external reminders to use messages or, in other cases, when events emerge, and the developers want to push recipients to use the messages, such as wearing masks during the COVID-19 pandemic. The signal can be an active prompt or cue in the form of an SMS text message delivered by developers [31,55]. Among the 71 included studies in the systematic review by Simeon et al [74], 10 studies reported the use of prompts or cues as a behavior change technique in self-directed social media programs. Webb et al [13] also found that SMS text messages were highly effective for behavior change in internet interventions when they provided cues to action. A signal, on the other hand, is an event that is emergent, accidental, or unexpected such as an adverse event that happened on a unit, a new health care regulation or policy, or a global pandemic. These events remind people of the relevant resources on social media platforms that can help tackle the situation. Together, the *SMILE* framework proposes sparks, facilitators, and signals as triggers to activate recipients to use social media messages.

Outcomes

In the *SMILE* framework, we specify the knowledge translation outcome as research use, which is a multidimensional concept that involves conceptual, instrumental, and persuasive use of research findings [33,34]. Conceptual research use refers to using research evidence to change the levels of knowledge, understanding, or attitude of a recipient. Both the *It Doesn't Have to Hurt* initiative [10,26] and *Be Sweet to Babies* initiative [27,35-39] have demonstrated that when recipients receive relevant and appropriate messages on social media that respond to their needs and context, they are highly likely to improve conceptual research use. Instrumental research use involves the direct application of research evidence in practice to change behavior. Modanloo et al [41] and Tunnecliff et al [76] have shown, in their randomized controlled trials, that different types of triggers, such as sparks or facilitators, are essential for the active uptake of research evidence and behavior change by recipients. Persuasive research use refers to using research evidence as a political or persuasive tool to justify an action, attain power, or achieve goals [33,34]. One of the most typical examples is the #WearingMasks social media campaign during COVID-19, which has made a huge impact on public behavior and government policy making.

Discussion

Principal Findings

In this paper, we present the *SMILE* framework, which is based on a review of 5 social media initiatives, 5 theories, and 58 empirical studies. The framework provides a preliminary understanding of how social media works as a knowledge translation strategy for health care providers, policy makers, and patients to inform their health care decision-making.

The *SMILE* framework has implications for research by offering a heuristic device for the development of social media interventions to promote evidence use. We suggest that it be used in combination with process frameworks, which provide step-by-step guidance on implementing web-based knowledge translation interventions [86] or evaluation frameworks to evaluate the multilevel outcomes and impacts of social media interventions [54].

Implications for Social Media Strategy Development

On the basis of this framework, we offer several suggestions for researchers and organizations who intend to use social media to promote research use. First, in the preparation stage, it is important for developers to assess their readiness to start a social media initiative. Some probing questions may be considered during this stage, such as is there an explicit topic to be covered? Does the team have enough time, resources, and expertise to develop the intervention and monitor the operation?

Once the infrastructure has been built for the social media initiative, the team begins developing a message and delivery strategy. Developers should recognize target users' needs and their context and, if possible, engage them in the development process. The six attributes of messages (ie, relevance, aesthetics, readability, findability, credibility, and usability) and three dimensions of delivery strategies (ie, social media platform, mode of delivery, and specific parameters) need to be taken into account when creating the social media interventions.

The team can then start the activation stage, where they make efforts to embed triggers into the social media delivery mechanisms for recipients to use the messages. Developers may interact with multilevel stakeholders and investigate the enablers of and barriers to recipients' use of the messages. By tailoring behavior change techniques to identified barriers and enablers, the development team can develop a social media strategy that has the greatest potential to affect message use in practice.

Acknowledgment of Complexity Within This Framework

We fully acknowledge the complexity of developing and implementing social media interventions and incorporate the notion of complexity within this framework in several ways. As information and communication sciences are fast-growing fields, new features and functions for social media platforms are continually emerging. Consequently, the approaches to developing messages and delivery strategies may become more diversified as technology advances. The dynamic interactions between constructs within the *SMILE* framework, such as the interaction between developers, recipients, and their situated contexts, make it challenging to undertake firm predictions [87,88]. Developers should immerse themselves in the human–social media system and capture underlining interactive patterns to inform the development of the most relevant and targeted activating techniques.

We also acknowledge the nonlinear aspect of the social media implementation processes, in which different levels of context influence and shape behavior. Promoting research use through social media is not a linear, straightforward process, and constant adaptations should be expected and embraced to optimize interventions. Finally, as each construct within this framework does not have a fixed and predetermined effect, and the interactions between constructs are dynamic and complex, we recognize that the framework has not been empirically validated and may not reveal all of the mechanisms at play for social media to influence research use. Nevertheless, the framework is based on current empirical evidence and well-recognized theories to provide plausible explanations for the successes and failures of social media interventions. Overall, the framework explicates the complexities of using social media in real-world practice and elucidates the key domains that developers, recipients, and researchers should attend to when developing or evaluating social media interventions.

As Maloney et al [72] suggested, "rather than looking at whether or not social media is effective for health professional education, it may be time to look at how various modalities can be optimized, both in terms of how the messages are delivered and how learners can be supported to engage." Using social media to disseminate research evidence has become such an inexorable global trend that researchers should go beyond the investigation of the effectiveness of social media interventions and delve into the theoretical field on how to make it effective. The next stage of our project will be to test and refine the *SMILE* framework through a realist methodology that unpacks the mechanisms of how and under what circumstances social media works as a knowledge translation strategy for health care professionals to improve the delivery of research-based care.

Limitations

The *SMILE* framework and its development process have 2 limitations. First, because of the multiple interactive components involved in developing and using social media for knowledge translation, as well as the massive amount of literature available from the relevant disciplinary fields, it was challenging to retrieve all pertinent theories and studies using a full systematic review approach. Instead, we used a targeted and flexible approach to select studies that allowed us to prioritize articles based on our framework's development needs. It is possible that we missed some research and embedded our own values into the propositions by using this approach; thus, our next step is to test and refine the framework. Second, as the use of social media for knowledge translation in real-world practice is still in its infancy, we could not locate studies that captured all the *SMILE* framework's propositions. More empirical studies of social media initiatives are needed to test the propositions of this framework.

Conclusions

In this paper, we propose the *SMILE* framework based on a review of social media initiatives, theories, and empirical studies as a preliminary understanding of how social media works as a knowledge translation strategy in health care decision-making. We provide a detailed description of each construct in the framework and offer suggestions for researchers and developers who intend to develop social media initiatives and interventions. For social media to be effective in enabling recipients to use research evidence in their practice decision-making, the *SMILE* framework purports that developers respond to target recipients' needs and context to develop relevant social media messages and appropriate delivery strategies. Recipients' use of messages is influenced by the virtual–technical, individual, organizational, and system contexts and can be activated by three types of triggers: sparks, facilitators, and signals. The *SMILE* framework maps the factors that are hypothesized to influence recipients' social media message use and offers a heuristic device for social media developers and researchers to develop social media interventions. More empirical studies and social media initiatives are needed to test the propositions of the *SMILE* framework.

Conflicts of Interest

None declared.

References

1. Sun M, Yang L, Chen W, Luo H, Zheng K, Zhang Y, et al. Current status of official WeChat accounts for public health education. J Public Health (Oxf) 2021 Sep 22;43(3):618-624. [doi: 10.1093/pubmed/fdz163] [Medline: 31974552]
2. Zhang X, Wen D, Liang J, Lei J. How the public uses social media wechat to obtain health information in China: a survey study. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):66 [FREE Full text] [doi: 10.1186/s12911-017-0470-0] [Medline: 28699549]
3. Warden C. Liquid Lock. URL: https://liquidlockmedia.com/30-social-media-and-healthcare-statistics/ [accessed 2022-02-01]
4. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
5. Hamm MP, Chisholm A, Shulhan J, Milne A, Scott SD, Given LM, et al. Social media use among patients and caregivers: a scoping review. BMJ Open 2013 May 09;3(5):e002819 [FREE Full text] [doi: 10.1136/bmjopen-2013-002819] [Medline: 23667163]
6. Dol J, Tutelman PR, Chambers CT, Barwick M, Drake EK, Parker JA, et al. Health researchers' use of social media: scoping review. J Med Internet Res 2019 Nov 13;21(11):e13687 [FREE Full text] [doi: 10.2196/13687] [Medline: 31719028]
7. Zhu Z, Xing W, Yan H, Zhou Y, Ying G, Cheng L. Construction and effect evaluation of platform for evidence dissemination. Chin J Nurs 2017;52(3):271-274. [doi: 10.3761/j.issn.0254-1769.2017.03.003]
8. Zhu Z, Xing W, Hu Y, Zhou Y, Gu Y. Improving evidence dissemination and accessibility through a mobile-based resource platform. J Med Syst 2018 May 28;42(7):118. [doi: 10.1007/s10916-018-0969-7] [Medline: 29808443]
9. Dyson MP, Newton AS, Shave K, Featherstone RM, Thomson D, Wingert A, et al. Social media for the dissemination of Cochrane child health evidence: evaluation study. J Med Internet Res 2017 Sep 01;19(9):e308 [FREE Full text] [doi: 10.2196/jmir.7819] [Medline: 28864427]
10. Chambers CT, Dol J, Parker JA, Caes L, Birnie KA, Taddio A, et al. Implementation effectiveness of a parent-directed YouTube video ("It doesn't have to hurt") on evidence-based strategies to manage needle pain: descriptive survey study. JMIR Pediatr Parent 2020 Mar 04;3(1):e13552 [FREE Full text] [doi: 10.2196/13552] [Medline: 32130190]
11. #ItDoesntHaveToHurt: making a difference for children---science-media partnership harnesses social media to connect with parents and mobilize evidence on children's pain. Canadian Institutes of Health Research. URL: http://cihr-irsc.gc.ca/e/51240.html [accessed 2022-02-01]
12. Evidence-based Mutual Aid Action Against Pandemic: You Ask, We Answer. Beijing University of Chinese Medicine Cochrane Center. URL: https://mp.weixin.qq.com/s/A8REQuPasFEwlpCn29IMwQ [accessed 2022-02-01]
13. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
14. Arguel A, Perez-Concha O, Li SY, Lau AY. Theoretical approaches of online social network interventions and implications for behavioral change: a systematic review. J Eval Clin Pract 2018 Feb;24(1):212-221. [doi: 10.1111/jep.12655] [Medline: 27709724]
15. Ngai EW, Tao SS, Moon KK. Social media research: theories, constructs, and conceptual frameworks. Int J Inf Manag 2015 Feb;35(1):33-44. [doi: 10.1016/j.ijinfomgt.2014.09.004]

XSL•FO
RenderX

16. Ngai EW, Moon KK, Lam S, Chin ES, Tao SS. Social media models, technologies, and applications: an academic review and case study. Industr Manag Data Syst 2015 Jun 08;115(5):769-802. [doi: 10.1108/IMDS-03-2015-0075]

17. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]

18. Ritterband LM, Bailey ET, Thorndike FP, Lord HR, Farrell-Carnahan L, Baum LD. Initial evaluation of an internet intervention to improve the sleep of cancer survivors with insomnia. Psychooncology 2012 Jul;21(7):695-705 [FREE Full text] [doi: 10.1002/pon.1969] [Medline: 21538678]

19. Corkum PV, Reid GJ, Hall WA, Godbout R, Stremler R, Weiss SK, et al. Evaluation of an internet-based behavioral intervention to improve psychosocial health outcomes in children with insomnia (better nights, better days): protocol for a randomized controlled trial. JMIR Res Protoc 2018 Mar 26;7(3):e76 [FREE Full text] [doi: 10.2196/resprot.8348] [Medline: 29581089]

20. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. Business Horizons 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]

21. Petkovic J, Duench S, Trawin J, Dewidar O, Pardo Pardo J, Simeon R, et al. Behavioural interventions delivered through interactive social media for health behaviour change, health outcomes, and health equity in the adult population. Cochrane Database Syst Rev 2021 May 31;5:CD012932. [doi: 10.1002/14651858.CD012932.pub2] [Medline: 34057201]

22. Shaw C. Implementing an online social network for health communication. The University of New Mexico UNM Digital Repository. 2014. URL: https://digitalrepository.unm.edu/cgi/viewcontent.cgi?article=1084&context=biom_etds [accessed 2022-02-01]

23. Meleis A. Theoretical Nursing: Development and Progress. Philadelphia, Pennsylvania, United States: Lippincott Williams & Wilkins; 2011.

24. Zhu Z, Xing W, Hu Y, Zhou Y, Gu Y, Cheng L. The construction of evidence dissemination platform based on mobile terminals and the status of the consumers' demand and experience. J Nurs Train 2017;32(10):939-941. [doi: 10.16821/j.cnki.hsjx.2017.10.029]

25. Using social media platforms to disseminate Cochrane evidence in China. Cochrane Community. URL: https://community.cochrane.org/news/using-social-media-platforms-disseminate-cochrane-evidence-china [accessed 2022-02-01]

26. Campbell-Yeo M, Dol J, Disher T, Benoit B, Chambers CT, Sheffield K, et al. The power of a parent's touch: evaluation of reach and impact of a targeted evidence-based YouTube video. J Perinat Neonatal Nurs 2017;31(4):341-349. [doi: 10.1097/JPN.0000000000000263] [Medline: 28520656]

27. Harrison D, Wilding J, Bowman A, Fuller A, Nicholls SG, Pound CM, et al. Using YouTube to disseminate effective vaccination pain treatment for babies. PLoS One 2016;11(10):e0164123 [FREE Full text] [doi: 10.1371/journal.pone.0164123] [Medline: 27695054]

28. Translating evidence in child health to enhance outcomes. EchoKT. URL: http://www.echokt.ca/ [accessed 2022-02-01]

29. Harvey G, Kitson A. PARIHS revisited: from heuristic to integrated framework for the successful implementation of knowledge into practice. Implement Sci 2016 Mar 10;11:33 [FREE Full text] [doi: 10.1186/s13012-016-0398-2] [Medline: 27013464]

30. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]

31. Fogg B. A behavior model for persuasive design. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009 Presented at: Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology; Apr 26 - 29, 2009; Claremont California USA. [doi: 10.1145/1541948.1541999]

32. Rogers E. Diffusion of Innovations, 5th Edition. New York, United States: Simon & Schuster; 2003.

33. Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? J Contin Educ Health Prof 2006;26(1):13-24. [doi: 10.1002/chp.47] [Medline: 16557505]

34. Straus S, Tetroe J, Graham I. Knowledge Translation in Health Care: Moving from Evidence to Practice. Hoboken, New Jersey, United States: John Wiley & Sons; 2009.

35. Harrison D, Larocque C, Reszel J, Harrold J, Aubertin C. Be sweet to babies during painful procedures: a pilot evaluation of a parent-targeted video. Adv Neonatal Care 2017 Oct;17(5):372-380. [doi: 10.1097/ANC.0000000000000425] [Medline: 28885227]

36. Vieira AC, Bueno M, Harrison D. "Be sweet to babies": use of Facebook as a method of knowledge dissemination and data collection in the reduction of neonatal pain. Paediatric Neo Pain 2020 May 02;2(3):93-100. [doi: 10.1002/pne2.12022]

37. Bueno M, Costa RN, de Camargo PP, Costa T, Harrison D. Evaluation of a parent-targeted video in Portuguese to improve pain management practices in neonates. J Clin Nurs 2018 Mar;27(5-6):1153-1159. [doi: 10.1111/jocn.14147] [Medline: 29076203]

38. Almeida HC, Candido LK, Harrison D, Bueno M. Be Sweet to Babies: evaluation of an instructional video on neonatal pain management by nurses. Rev Esc Enferm USP 2018 Jun 25;52:e03313 [FREE Full text] [doi: 10.1590/S1980-220X2017033903313] [Medline: 29947704]

39. Hu J, Gifford W, Zhou Y, Zhang Q, Harrison D. Nurses' perspectives on pain management practices during newborn blood sampling in China. J Neonatal Nurs 2021 Dec;27(6):483-487. [doi: 10.1016/j.jnn.2021.06.015]

40. Hu J, Ruan H, Li Q, Gifford W, Zhou Y, Yu L, et al. Barriers and facilitators to effective procedural pain treatments for pediatric patients in the Chinese context: a qualitative descriptive study. J Pediatr Nurs 2020;54:78-85. [doi: 10.1016/j.pedn.2020.06.004] [Medline: 32585541]

41. Modanloo S, Dunn S, Stacey D, Harrison D. The feasibility, acceptability and preliminary efficacy of parent-targeted interventions in vaccination pain management of infants: a pilot randomized control trial (RCT). Pain Manag 2021 May;11(3):287-301. [doi: 10.2217/pmt-2020-0072] [Medline: 33593096]

42. Korki de Candido L, Harrison D, Ramallo Veríssimo MD, Bueno M. Effectiveness of a parent‐targeted video on neonatal pain management: nonrandomized pragmatic trial. Paediatric Neo Pain 2020 May 06;2(3):74-81. [doi: 10.1002/pne2.12023]

43. Reid K, Hartling L, Ali S, Le A, Norris A, Scott SD. Development and usability evaluation of an art and narrative-based knowledge translation tool for parents with a child with pediatric chronic pain: multi-method study. J Med Internet Res 2017 Dec 14;19(12):e412 [FREE Full text] [doi: 10.2196/jmir.8877] [Medline: 29242180]

44. Abdel-Wahab N, Rai D, Siddhanamatha H, Dodeja A, Suarez-Almazor ME, Lopez-Olivo MA. A comprehensive scoping review to identify standards for the development of health information resources on the internet. PLoS One 2019;14(6):e0218342 [FREE Full text] [doi: 10.1371/journal.pone.0218342] [Medline: 31220126]

45. Chang J, Rosenn I, Backstrom L, Marlow C. ePluribus: ethnicity on social networks. In: Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media. 2010 Presented at: Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media; May 23 – 26, 2010; Washington, D.C.

46. Correa T, Hinsley AW, de Zúñiga HG. Who interacts on the web?: the intersection of users' personality and social media use. Comput Human Behav 2010 Mar;26(2):247-253. [doi: 10.1016/j.chb.2009.09.003]

47. Dubowicz A, Schulz PJ. Medical information on the internet: a tool for measuring consumer perception of quality aspects. Interact J Med Res 2015 Mar 30;4(1):e8 [FREE Full text] [doi: 10.2196/ijmr.3144] [Medline: 25835333]

48. Rauniar R, Rawski G, Yang J, Johnson B. Technology acceptance model (TAM) and social media usage: an empirical study on Facebook. J Ent Info Management 2014 Feb 04;27(1):6-30. [doi: 10.1108/JEIM-04-2012-0011]

49. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]

50. Hazzam J, Lahrech A. Health care professionals' social media behavior and the underlying factors of social media adoption and use: quantitative study. J Med Internet Res 2018 Nov 07;20(11):e12035 [FREE Full text] [doi: 10.2196/12035] [Medline: 30404773]

51. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013 Jan;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]

52. McInnes N, Haglund BJ. Readability of online health information: implications for health literacy. Inform Health Soc Care 2011 Dec;36(4):173-189. [doi: 10.3109/17538157.2010.542529] [Medline: 21332302]

53. Mislove A, Lehmann S, Ahn Y, Onnela J, Rosenquist J. Understanding the demographics of Twitter users. In: Proceedings of the Fifth International Conference on Weblogs and Social Media. 2011 Presented at: Proceedings of the Fifth International Conference on Weblogs and Social Media; Jul 17-21, 2011; Barcelona, Catalonia, Spain.

54. O'Grady L, Witteman H, Bender JL, Urowitz S, Wiljer D, Jadad AR. Measuring the impact of a moving target: towards a dynamic framework for evaluating collaborative adaptive interactive technologies. J Med Internet Res 2009 Jun 18;11(2):e20 [FREE Full text] [doi: 10.2196/jmir.1058] [Medline: 19632973]

55. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009 Mar;24(1). [doi: 10.17705/1CAIS.02428]

56. Robins D, Holmes J. Aesthetics and credibility in web site design. Inf Process Manag 2008 Jan;44(1):386-399. [doi: 10.1016/j.ipm.2007.02.003]

57. Schein R, Wilson K, Keelan J. Literature review on effectiveness of the use of social media: a report for Peel public health. Region of Peel. URL: https://peelregion.ca/health/resources/pdf/socialmedia.pdf [accessed 2022-02-10]

58. Schubart JR, Stuckey HL, Ganeshamoorthy A, Sciamanna CN. Chronic health conditions and internet behavioral interventions: a review of factors to enhance user engagement. Comput Inform Nurs 2011 Feb;29(2):81-92 [FREE Full text] [doi: 10.1097/NCN.0b013e3182065eed] [Medline: 21164337]

59. Short CE, Gelder C, Binneweg L, McIntosh M, Turnbull D. Examining the accessibility of high-quality physical activity behaviour change support freely available online for men with prostate cancer. J Cancer Surviv 2018 Feb;12(1):10-17. [doi: 10.1007/s11764-017-0638-8] [Medline: 28871558]

60. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the internet: caveat lector et viewor--Let the reader and viewer beware. JAMA 1997 Apr 16;277(15):1244-1245. [Medline: 9103351]

61. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]

62. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. J Med Internet Res 2019 May 02;21(5):e12522 [FREE Full text] [doi: 10.2196/12522] [Medline: 31045507]

63. Zhang Y, Sun Y, Xie B. Quality of health information for consumers on the web: a systematic review of indicators, criteria, tools, and evaluation results. J Assn Inf Sci Tec 2015 Apr 29;66(10):2071-2084. [doi: 10.1002/asi.23311]

64. Zhong B, Hardin M, Sun T. Less effortful thinking leads to more social networking? The associations between the use of social network sites and personality traits. Comput Human Behav 2011 May;27(3):1265-1271. [doi: 10.1016/j.chb.2011.01.008]

65. Zolkepli IA, Kamarulzaman Y. Social media adoption: the role of media needs and innovation characteristics. Comput Human Behav 2015 Feb;43:189-209. [doi: 10.1016/j.chb.2014.10.050]

66. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011 Jan 06;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]

67. Brouwer W, Oenema A, Crutzen R, de Nooijer J, de Vries NK, Brug J. An exploration of factors related to dissemination of and exposure to internet-delivered behavior change interventions aimed at adults: a Delphi study approach. J Med Internet Res 2008 Apr 16;10(2):e10 [FREE Full text] [doi: 10.2196/jmir.956] [Medline: 18417443]

68. Brouwer W, Oenema A, Crutzen R, de Nooijer J, de Vries N, Brug J. What makes people decide to visit and use an internet‐delivered behavior‐change intervention? Health Education 2009 Oct 16;109(6):460-473. [doi: 10.1108/09654280911001149]

69. Dawot N, Ibrahim R. A review of features and functional building blocks of social media. In: Proceedings of the 2014 8th Malaysian Software Engineering Conference (MySEC). 2014 Presented at: 2014 8th. Malaysian Software Engineering Conference (MySEC); Sep 23-24, 2014; Langkawi, Malaysia. [doi: 10.1109/mysec.2014.6986010]

70. Elaheebocus SM, Weal M, Morrison L, Yardley L. Peer-based social media features in behavior change interventions: systematic review. J Med Internet Res 2018 Feb 22;20(2):e20 [FREE Full text] [doi: 10.2196/jmir.8342] [Medline: 29472174]

71. Ko N, Hsieh C, Wang M, Lee C, Chen C, Chung A, et al. Effects of internet popular opinion leaders (iPOL) among Internet-using men who have sex with men. J Med Internet Res 2013 Feb 25;15(2):e40 [FREE Full text] [doi: 10.2196/jmir.2264] [Medline: 23439583]

72. Maloney S, Tunnecliff J, Morgan P, Gaida JE, Clearihan L, Sadasivan S, et al. Translating evidence into practice via social media: a mixed-methods study. J Med Internet Res 2015 Oct 26;17(10):e242 [FREE Full text] [doi: 10.2196/jmir.4763] [Medline: 26503129]

73. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]

74. Simeon R, Dewidar O, Trawin J, Duench S, Manson H, Pardo Pardo J, et al. Behavior change techniques included in reports of social media interventions for promoting health behaviors in adults: content analysis within a systematic review. J Med Internet Res 2020 Jun 11;22(6):e16002 [FREE Full text] [doi: 10.2196/16002] [Medline: 32525482]

75. Tajudeen FP, Jaafar NI, Ainin S. Understanding the impact of social media usage among organizations. Inf Manag 2018 Apr;55(3):308-321. [doi: 10.1016/j.im.2017.08.004]

76. Tunnecliff J, Weiner J, Gaida JE, Keating JL, Morgan P, Ilic D, et al. Translating evidence to practice in the health professions: a randomized trial of Twitter vs Facebook. J Am Med Inform Assoc 2017 Mar 01;24(2):403-408 [FREE Full text] [doi: 10.1093/jamia/ocw085] [Medline: 27357833]

77. Wirtz B, Piehler R, Ullrich S. Determinants of social media website attractiveness. J Electronic Commerce Res 2013 Feb;14(1):11-33.

78. Barton C. The current sports medicine journal model is outdated and ineffective. Aspetar Sports Med J 2017;6:58-63.

79. Gates A, Featherstone R, Shave K, Scott SD, Hartling L. Dissemination of evidence in paediatric emergency medicine: a quantitative descriptive evaluation of a 16-week social media promotion. BMJ Open 2018 Jun 06;8(6):e022298 [FREE Full text] [doi: 10.1136/bmjopen-2018-022298] [Medline: 29880576]

80. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]

81. Guidance for the Assessment of Context and Implementation in Health Technology Assessments (HTA) and Systematic Reviews of Complex Interventions: The Context and Implementation of Complex Interventions (CICI) Framework. -: Integrated Health Technology Assessment for Evaluating Complex Technologies (INTEGRATE-HTA); 2016.

82. Chapman SJ, Grossman RC, FitzPatrick ME, Brady RR. Randomized controlled trial of plain English and visual abstracts for disseminating surgical research via social media. Br J Surg 2019 Oct 02:1611-1616. [doi: 10.1002/bjs.11307] [Medline: 31577372]

83. Ibrahim AM, Lillemoe KD, Klingensmith ME, Dimick JB. Visual abstracts to disseminate research on social media: a prospective, case-control crossover study. Ann Surg 2017 Dec;266(6):e46-e48. [doi: 10.1097/SLA.0000000000002277] [Medline: 28448382]

84. Fudan University Evidence-Based Nursing Center Anual Report. Yunzhan. URL: http://www.yunzhan365.com/homepage/dojt [accessed 2022-02-01]

85. It doesn't have to hurt team. Dr. Christine Chambers. URL: https://itdoesnthavetohurt.ca/team/#team1 [accessed 2022-02-01]

86. Levac D, Glegg SM, Camden C, Rivard LM, Missiuna C. Best practice recommendations for the development, implementation, and evaluation of online knowledge translation resources in rehabilitation. Phys Ther 2015 Apr;95(4):648-662. [doi: 10.2522/ptj.20130500] [Medline: 25301966]

87. Kitson A, Brook A, Harvey G, Jordan Z, Marshall R, O'Shea R, et al. Using complexity and network concepts to inform healthcare knowledge translation. Int J Health Policy Manag 2018 Mar 01;7(3):231-243 [FREE Full text] [doi: 10.15171/ijhpm.2017.79] [Medline: 29524952]

88. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]

Abbreviations

BUCM: Beijing University of Chinese Medicine
COM-B: capability, opportunity, motivation, and behavior
ECHO: Evidence in Child Health to Enhance Outcomes
i-PARIHS: integrated Promoting Action on Research Implementation in Health Services
JBI: Joanna Briggs Institute
SMILE: Social Media for Implementing Evidence

<u>Original Paper</u>

Optimizing Existing Mental Health Screening Methods in a Dementia Screening and Risk Factor App: Observational Machine Learning Study

Narayan Kuleindiren[1*], MSci; Raphael Paul Rifkin-Zybutz[1*], BA, BM BCh; Monika Johal[1,2], BSc; Hamzah Selim[1,3], BSc; Itai Palmon[1,4], BSc; Aaron Lin[1,5]; Yizhou Yu[1], BSc; Ali Alim-Marvasti[1,6], BA, MA, MBBChir, MRCP; Mohammad Mahmud[1,7], BSc, MBBS, MRCP(UK)

[1]Mindset Technologies Ltd, London, United Kingdom

[2]Imperial College School of Medicine, Faculty of Medicine, Imperial College London, London, United Kingdom

[3]Medical School, University College London, London, United Kingdom

[4]Medical School, University of Michigan, Ann Arbor, MI, United States

[5]Medical School, University of Birmingham, Birmingham, United Kingdom

[6]Queen Square Institute of Neurology, University College London, London, United Kingdom

[7]Department of Brain Sciences, Imperial College London, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Raphael Paul Rifkin-Zybutz, BA, BM BCh
Mindset Technologies Ltd
3rd Floor, 5 Chancery Lane
London, WC2A 1LG
United Kingdom
Phone: 44 02081522341
Email: <u>raphael.rifkinzybutz@meetmindset.com</u>

Abstract

Background: Mindstep is an app that aims to improve dementia screening by assessing cognition and risk factors. It considers important clinical risk factors, including prodromal symptoms, mental health disorders, and differential diagnoses of dementia. The 9-item Patient Health Questionnaire for depression (PHQ-9) and the 7-item Generalized Anxiety Disorder Scale (GAD-7) are widely validated and commonly used scales used in screening for depression and anxiety disorders, respectively. Shortened versions of both (PHQ-2/GAD-2) have been produced.

Objective: We sought to develop a method that maintained the brevity of these shorter questionnaires while maintaining the better precision of the original questionnaires.

Methods: Single questions were designed to encompass symptoms covered in the original questionnaires. Answers to these questions were combined with PHQ-2/GAD-2, and anonymized risk factors were collected by Mindset4Dementia from 2235 users. Machine learning models were trained to use these single questions in combination with data already collected by the app: age, response to a joke, and reporting of functional impairment to predict binary and continuous outcomes as measured using PHQ-9/GAD-7. Our model was developed with a training data set by using 10-fold cross-validation and a holdout testing data set and compared to results from using the shorter questionnaires (PHQ-2/GAD-2) alone to benchmark performance.

Results: We were able to achieve superior performance in predicting PHQ-9/GAD-7 screening cutoffs compared to PHQ-2 (difference in area under the curve 0.04, 95% CI 0.00-0.08, P=.02) but not GAD-2 (difference in area under the curve 0.00, 95% CI –0.02 to 0.03, P=.42). Regression models were able to accurately predict total questionnaire scores in PHQ-9 (R^2=0.655, mean absolute error=2.267) and GAD-7 (R^2=0.837, mean absolute error=1.780).

Conclusions: We app-adapted PHQ-4 by adding brief summary questions about factors normally covered in the longer questionnaires. We additionally trained machine learning models that used the wide range of additional information already collected in Mindstep to make a short app-based screening tool for affective disorders, which appears to have superior or equivalent performance to well-established methods.

XSL•FO

RenderX

Introduction

Depression was among the 12 modifiable dementia risk factors identified by the Lancet commission [1]. The relationship between depression and dementia is complex, with depression being a risk factor for dementia, a prodromal symptom [2] and a differential diagnosis known as pseudodementia [3]. Anxiety is a highly comorbid condition with depression and is an important feature in the diagnosis of dementia [4]. Anxiety has independent effects on cognition [5] and plays a role in driving health-seeking behavior in individuals without deficits [6]. Furthermore, depression and anxiety symptoms are common in older adults with an estimated prevalence of around 13% [7]. Therefore, appropriate screening for depression and anxiety is of importance when screening for dementia. Mindstep is a new app that aims to holistically screen for cognitive impairment and dementia while gathering information on important dementia risk factors. It achieves this by integrating the analyses of important risk factors such as depression and anxiety with cognitive screening tests in a conversation interface. It is important that the methods used within this app are accurate and easy to integrate within the wider structure of the app.

Both the 9-item Patient Health Questionnaire for depression (PHQ-9) and the 15-item Geriatric Depression Scale are widely used in clinical settings to screen for depression in older adults [8,9]. However, despite being shorter, PHQ-9 performs at least as well as the Geriatric Depression Scale in screening older adults across multiple populations, and therefore, we decided to incorporate PHQ-9 into the app [10,11]. The optimal cutoff point for the diagnosis of depression in PHQ-9 is ≥10 with an associated sensitivity and specificity of 88% [8]. Similarly, the 7-item Generalized Anxiety Disorder Scale (GAD-7) assessment has been used for anxiety screening with high sensitivity (92%) and specificity (76%) in those of working age with cutoff points ranging from ≥7 to ≥10 [12-14]. Further research has established effectiveness in screening older adults, with lower cutoffs of 5 recommended for better sensitivity [12,13].

Although these questionnaires are individually brief, they can become lengthy when nested within an app that aims to screen for multiple other risk factors and utilize multiple tests. Longer questionnaires cause higher rates of fatigue and dropouts, and hence, we aimed to limit the duration of total software use to 5 minutes [15]. This is especially key for individuals with affective disorders who are likely to experience deficits in attention, concentration, motivation, and fatigue [16,17]. We therefore considered PHQ-4 that combines PHQ-2 (which consists of the first 2 questions of PHQ-9) and GAD-2 (which consists of the first two questions of GAD-7) [18]. Although this shortens the time spent on the questionnaires, PHQ-4 does not have a severity scale, and commonly used cutoffs can result in prioritizing sensitivity or specificity at the expense of the other [19]. Consequently, this requires follow-up questions; for example,

completing the whole PHQ-9 following a positive screen on PHQ-2 [20]. We wanted to develop a method that would have both the brevity of PHQ-4 and the accuracy of the longer PHQ-9/GAD-7 in addition to a severity scale. To achieve this, we adapted PHQ-4 by adding questions about factors normally covered in the longer questionnaires. We trained machine learning models that used the wide range of additional information already collected in Mindstep. Therefore, this study aims to assess the performance of our models when benchmarked against full-length standardized PHQ-9 and GAD-7 questionnaires. If performance was equivalent to that of these longer questionnaires, this would enable the app to have a screening method of equivalent efficacy while minimizing completion time.

Methods

App Data Collection and Users

Data of users of the Mindstep app were used for this study in a convenience sample. The app consists of a 5-minute conversational style questionnaire where information on common dementia risk factors is gathered. Cognitive performance is assessed through modified versions of 2 common cognitive tests: the Stroop test and Symbols Digit Modalities test. The dementia risk factors queried include medical history, age, alcohol consumption and dependency, concussion, smoking, and self-reported functional impairment. Analogous to medical consultation, further screening is then performed in response to answers that would elicit concern. Only those who reported feeling depressed or tired were screened with PHQ-9, while only those who reported feeling anxious or worried were screened with GAD-7. As a control group, for a short time, those who answered that they felt fine or great were also screened for anxiety. Apart from age, no other personal information was gathered from users.

New Question Design

New questions were created based on the longer versions of PHQ-9 and GAD-7, each for depression (Mindset Depression Question [MDQ]) and anxiety (Mindset Anxiety Question [MAQ]) to encompass symptoms of depression and anxiety in the Diagnostic and Statistical Manual of Mental Disorders that would normally be excluded from the shortened questionnaires. In 1 question, users were asked to select in a binary manner if any of several options applied to them—a method of collecting a wide range of information in a rapid manner. The questions are shown below (Figure 1). A mixture of categorical and continuous features from the app was used. The features were selected by unsupervised recursive feature elimination. Both models included age, alcohol dependency (as assessed by CAGE) [21], and a functional impairment question. For the PHQ models, the MDQ, PHQ-2, alcohol/drugs/smoking (currently/past/never), weekly alcohol consumption in drinks

(<3/4-7/7-14/>14), feeling (depressed/tired), and joke data (yes/no/didn't get it) were included. For the GAD models, MAQ, GAD-2, and the cognitive scores (MStroop/MSymbols) were included.

Figure 1. (A) Screenshot of the Mindset Anxiety Question. (B) Screenshot of the Mindset Depression Question.
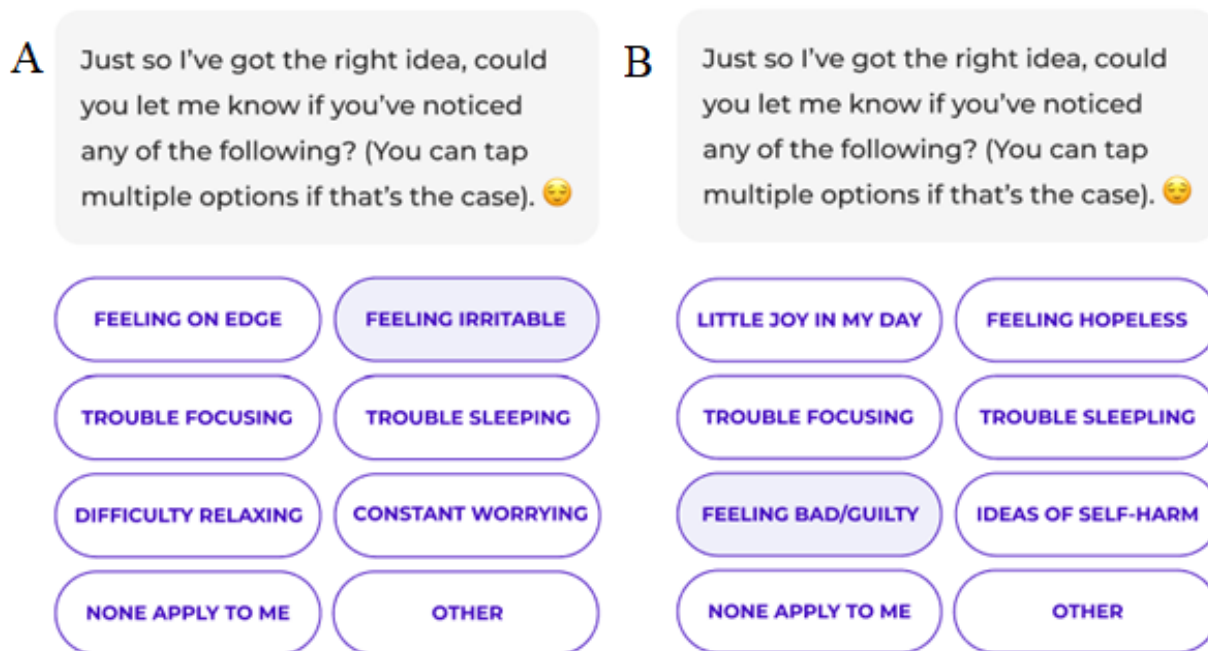
Benchmarks

Outcomes of interest were compared to those of the full-length PHQ-9 and GAD-7. For the binary classification task, a cutoff value of ≥10 was used in both cases to represent commonly used screening cutoffs for these tools [14,22]. The total PHQ-9 and GAD-7 scores were also used for the regression task where the full-length screening score was predicted from PHQ-4 plus Mindset features.

Data Preprocessing and Analysis

As part of the preprocessing pipeline, the categorical inputs were one-hot encoded. As only self-reported heavy drinkers were asked to complete the CAGE questionnaire, most users did not register a CAGE score and their score was assumed to be 0. For users who chose not to complete the Stroop or Symbols test, a mean value was used for their results. We divided the data into 80% training and a 20% holdout testing sets for PHQ (n=432,108) and GAD (n=408,103) with 10-fold cross validation. The holdout test set remained unseen throughout model training, hyperparameter tuning, and model selection.

Machine Learning Models

Four distinct machine learning models were trained in both classification and regression task. The models used were logistic regression (linear regression was used in the regression task), support vector machines, TabNet [23], and extreme gradient boosted trees. The models were evaluated based on 10-fold cross validation scores (area under the curve [AUC] for classification, R^2 for regression). The weaker models were discarded (AUC<0.9 or R^2<0.7), and the final result was a median ensemble of the remaining models. Only the ensemble model was tested on the testing set. The performance on the training set of this ensemble model compared to that on the individual models can be found in Table S1 of Multimedia Appendix 1.

Benchmarking

To guide interpretation and benchmarking of results, the final ensemble model was compared to logistic/linear regression models built using only the PHQ-2 and GAD-2 questionnaires. This was done on the unseen holdout test set. Confidence intervals and *P* values were generated to assess for the significance of differences by comparing model performance via a 1000-times bootstrap of the test set.

Model Explainability

To interpret the predictions of the final ensemble model, model agnostic Shapley Additive Explanations (SHAP) scores were calculated to determine the relative feature importance [24]. SHAP scores determine which features are important to the model across the entire testing set and enable local interpretations such as why a particular prediction was made for a given user.

Ethical Approval

This paper was a secondary data analysis of robustly anonymized data with minimal demographic information collected (only age) where there is no chance of data being linked to any individuals. On using the app, users agreed to transparent terms and conditions, which included having their data stored and anonymously used for further research. Therefore, ethical approval was not strictly required for this research. Out of an abundance of caution, we applied for and

were granted retrospective ethical approval for the use of these data for research: West Midlands, Solihull Research Ethics Committee, Reference 21/WM/0202.

Results

Participant Data

Of the 2235 Mindstep users, 540 completed the PHQ and 511 completed the GAD. The mean age of the total Mindstep users was 50 (SD 14.1) years; for the PHQ subset, it was 49.3 (SD 13.1) years; and for the GAD users, it was 49.1 (SD 13.6) years. Of the 540 targeted users selected to take PHQ-9, 233 (43.1%) screened positive for depression. Of the 511 targeted users selected to take the GAD-7, 173 (33.9%) screened positive for GAD. These high rates likely represent the enriched selection of users who had already reported feeling negatively valenced emotions. Only a small number of users did not complete the Stroop (28/2235, 1.3%) or Symbols tests (26/2235, 1.2%) and had data imputed.

Questionnaire Characteristics

All questionnaires had excellent reliability as measured by the Cronbach interitem correlation [25]: PHQ-9 (α=.84), PHQ-2 (α=.77), GAD-7 (α=.90), and GAD-2 (α=.84). The test set AUC for the ensemble model for PHQ-9 (0.90) was a significant improvement on the PHQ-2 baseline (difference 0.04, 95% CI 0.00-0.08; P=.02). The test set AUC for the ensemble model for GAD-7 (0.93) was equivalent to the GAD-2 baseline (difference 0.00, 95% CI –0.02 to 0.03, P=.42) (Figure 2). By altering thresholds, the sensitivity and specificity of the ensemble models can be optimized for particular situations. The selected optimal sensitivity and specificity for the PHQ model was 88% and 85%, respectively, achieving a good compromise compared to the highly sensitive PHQ-2 cutoff ≥2 (90% and 61%, respectively) or highly specific PHQ-2 cutoff ≥3 (68% and 93%, respectively). The positive and negative predictive values for the PHQ model were 78% and 92%, respectively. The sensitivity and specificity of the GAD model (95% and 78%, respectively) were substantively similar to the clinically used GAD-2 cutoff ≥3 (92% and 75%, respectively). The positive and negative predictive values for the GAD model were 69% and 96%, respectively.

Figure 2. Receiver operating characteristic curve for prediction of (A) Patient Health Questionnaire for depression and (B) Generalized Anxiety Disorder Scale of the model ensemble and their respective baselines. GAD: Generalized Anxiety Disorder; PHQ: Patient Health Questionnaire; ROC: receiver operating characteristic.
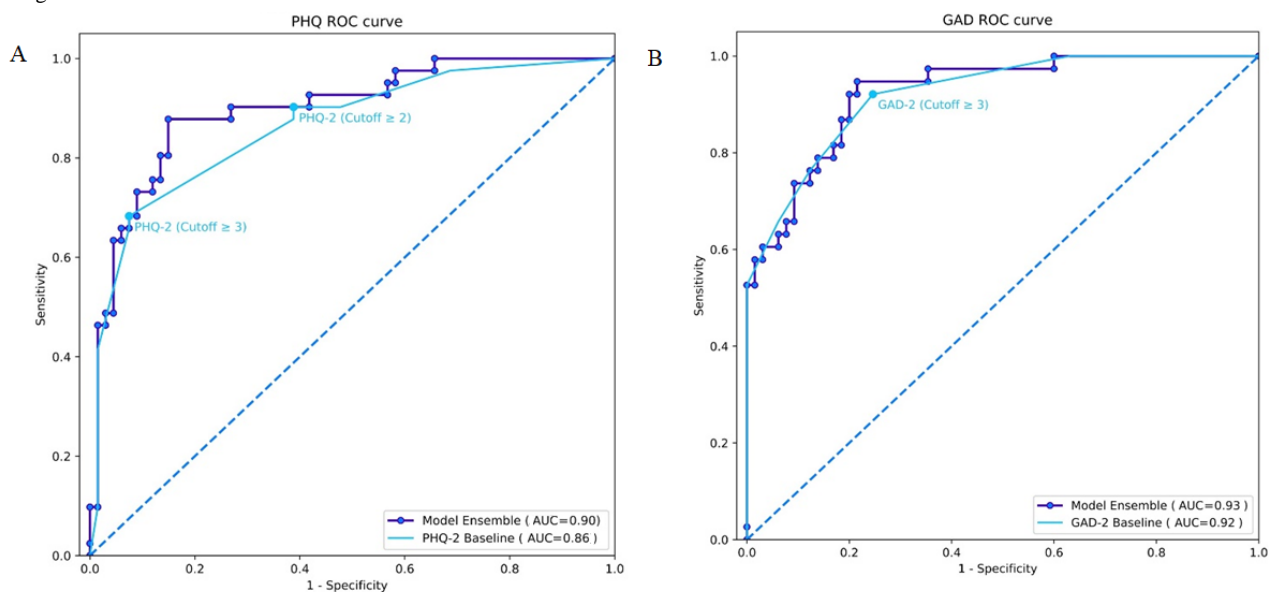
Regression Analysis

Figure 3 shows the regression model ensemble predictions for the test set PHQ-9 and GAD-7 scores. Both models (Figure 3) were able to achieve good prediction of the full-length questionnaire scores: PHQ-9 (R^2=0.655, mean absolute error [MAE]=2.267) and GAD-7 (R^2=0.837, MAE=1.780). The PHQ model showed a significant improvement in MAE over the PHQ-2 baseline (difference 0.35, 95% CI 0.06-0.65, P=.01) and a nonsignificant improvement in R^2 (0.08, 95% CI –0.02 to 0.21, P=.06). The GAD model showed a nonsignificant improvement in MAE over the GAD-2 baseline (0.08, 95% CI –0.10 to 0.26, P=.20) and a significant improvement in R^2 (0.04, 95% CI 0.01-0.08, P=.01) (Table S1 in Multimedia Appendix 1). For PHQ, by breaking the scores into categories of increasing severity [8], 0-4, 5-9, 10-14, 15-19, and ≥20, an intraclass correlation was calculated to be 0.76 (95% CI 0.67-0.83, P<.001). For GAD, categories of 0-4, 5-9, 10-14, and ≥15 were used [14]. An intraclass correlation was calculated as 0.87 (95% CI 0.81-0.91, P<.001).

Figure 3. The regression model ensemble predictions for the test set (A) 9-item Patient Health Questionnaire for depression and (B) 7-item Generalized Anxiety Disorder scale scores. GAD-7: 7-item Generalized Anxiety Disorder scale; MAE: mean absolute error; PHQ-9: 9-item Patient Health Questionnaire for depression.
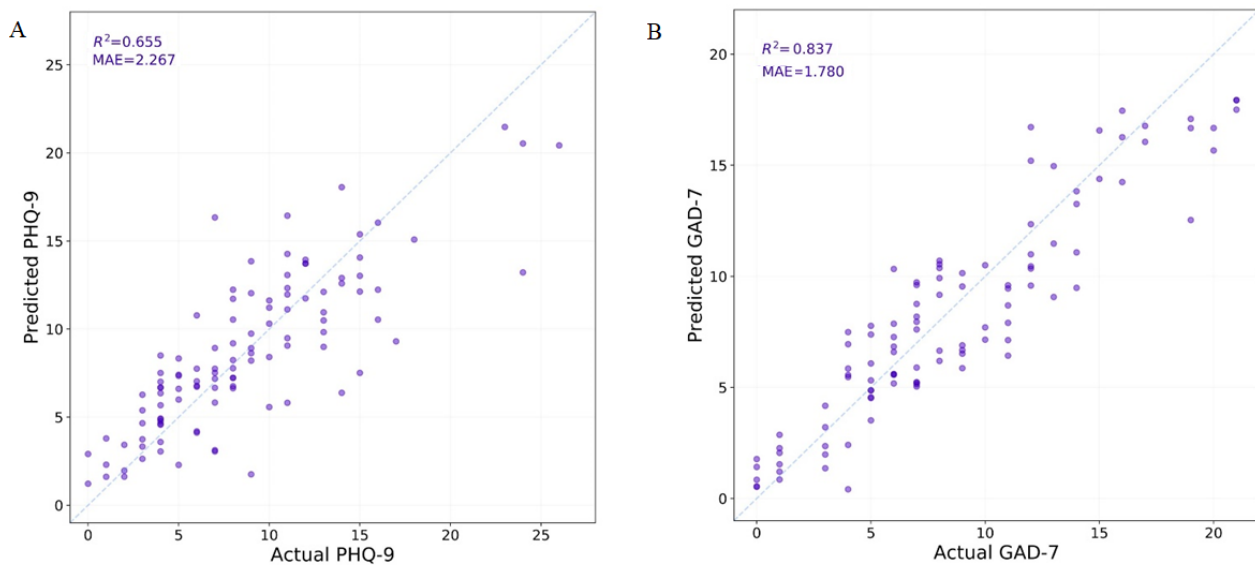
Feature Importance

Figure 4 shows the beeswarm SHAP summary plots of the 10 most important features as determined by SHAP values for predicting the PHQ-9 and GAD-7. The greater the magnitude of the SHAP values, the larger was the influence on the model with positive numbers, indicating that the user is more likely

to have the condition. In both cases, the most important features in the prediction of the full-length questionnaires were the PHQ-2 and GAD-2 followed by the MDQ and MAQ, respectively. The functional impairment question was also shown to be important in both sets of models. Age and smoking were important in the depression models, and the Stroop test was important in the anxiety models.

Figure 4. Beeswarm Shapley Additive Explanations summary plots of the 10 most important features as determined by Shapley Additive Explanations values for predicting the (A) 9-item Patient Health Questionnaire for depression and (B) 7-item Generalized Anxiety Disorder scale. Ordered by importance (top to bottom). The greater the magnitude of the Shapley Additive Explanations values, the larger the influence on the model with positive numbers, indicating the user is more likely to have the condition. FIQ: functional impairment question; GAD-7: 7-item Generalized Anxiety Disorder scale; MAQ: Mindset Anxiety Question; MDQ: Mindset Depression Question; PHQ-9: 9-item Patient Health Questionnaire for depression; SHAP: Shapley Additive Explanations.
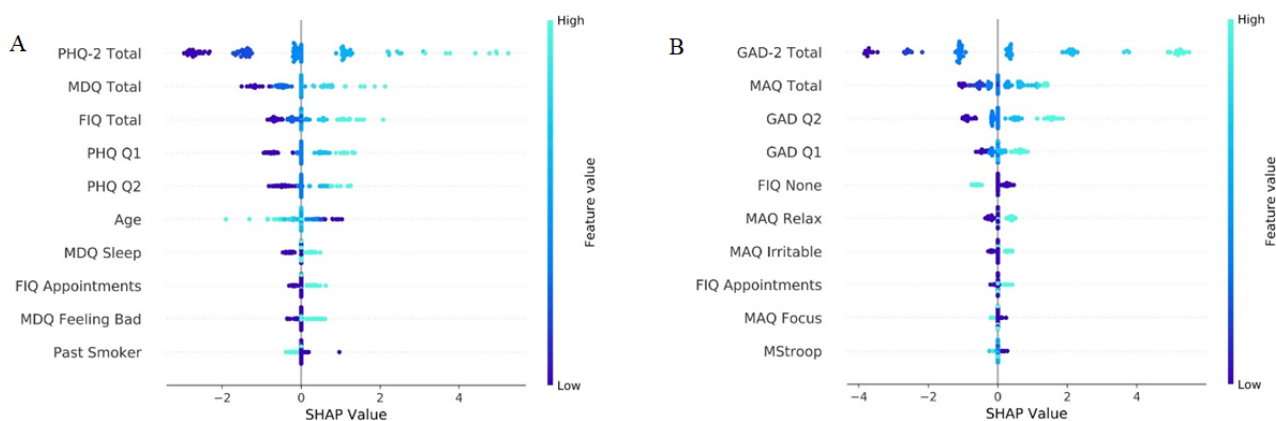
Discussion

Principal Results

We have shown that the combination of PHQ-4, additional questions, and risk factor information are able to accurately predict the severity score of the longer questions with an R^2 of 0.655 on PHQ-9 and R^2 of 0.837 for GAD-7. This is a novel finding, as previous studies have only looked at agreement

between the binary cutoffs of the shorter questionnaires compared to that of the longer. This suggests that even these ultrashort questionnaires may be responsive to change, although this will need to be explored in future work. In addition to this, compared to using the PHQ-2 alone, our model achieves significantly better performance on both classification and regression models. The SHAP analysis suggests that the MDQ and MAQ can capture some of the variance missed by the shorter PHQ-4. The benefit of our model is less clear in anxiety

with little difference compared to utilizing the GAD-2 alone. This may be in part, as GAD-2 alone achieves very high performance on both the binary classification and regression task. This is in line with meta-analyses, which show that GAD-2 achieves very similar performance to GAD-7 [12-14]. The strength of our developed model for depression is not just an enhanced accuracy of prediction but also our ability to choose any threshold to best balance sensitivity and specificity. This enables us to choose a cutoff that best balances sensitivity and specificity rather than having to choose between a PHQ-2 cutoff, which prioritizes a high sensitivity or specificity. Furthermore, the fact that we collect age will enable us to personalize the cutoff for screening in line with evidence by using less stringent cutoffs in older adults to maximize sensitivity [12,13]. Our regression model with good intraclass correlations enables us to sort users into multiple categories. For example, initial validation studies of PHQ-9 demonstrated that while 10 represented the best cutoff for sensitivity and specificity, higher scores had much better discriminative values with scores above 15 highly specific for depressive disorders and 10-15 representing an important grey zone [8]. Therefore, sorting users in multiple categories such as unlikely (<10)/possible (10-15)/probable (>15) depression is achievable and will assist in optimizing the accuracy of advice we can offer. Data suggest that longer questionnaires incur more fatigue and dropout [15]. This is especially important, as our app is self-administered and there is no clinician to encourage the user. The benefit of using ultrashort questionnaires is that it allows for many different risk factors to all be assessed with a single app and in a single sitting without significant fatigue and dropout. This enables a comprehensive review of many risk factors for dementia.

Limitations

Although these questionnaires are filled out unsupervised on an app, the validity of the computerized forms of PHQ-9 and GAD-7 has been demonstrated to be valid across format types [26], and the excellent reliability achieved in this study negate this as an issue. A limitation of this study is the use of PHQ-9/GAD-7 as our ground truth. This is an indirect measure compared to clinician-assessed diagnosis or structured diagnostic questionnaires such as the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders-IV (SCID). This both means that we cannot train our model on a definite diagnosis of depression. However, previous studies that have looked for proxies of depression diagnosis achieved AUCs of 0.77 [26] and 0.79 [27], which are low compared to the accuracy of the full PHQ-9 (AUC 0.87) [28]. This suggests that PHQ-9 is a valid ground truth to use. An additional limitation of the use of PHQ-9/GAD-7 as our ground truth means that our model can never outperform these questionnaires. Considering that we collect data on a wider variety of factors including

functional impairment, it may be that comparison to diagnostic measures would allow enhanced performance. Indeed, by capturing functional measures, including cognitive performance and self-perceived deficits, our app captures important elements missing from the PHQ-9 and GAD-7 questionnaires. Interestingly, the SHAP analysis shows that the questions around functional deficits are especially important in the PHQ-9 model, suggesting this information is important to the model's outperformance of the PHQ-2 alone. The next step to address this limitation would be to use the app in a population with gold standard validated measures of mental illness such as a SCID conducted by a mental health professional. This would both allow refinement of factors in the model and allow assessment of true sensitivity and specificity when it comes to the diagnosis of depression. However, it is important to note that the intent of the app is to use this function strictly in a screening and not a diagnostic role, with identified individuals being signposted for further assessment by their primary care doctors. Therefore, correlation to an already extremely well-validated questionnaire is likely to be adequate for its purpose.

Owing to the need for the app to collect anonymized data, we do not have basic demographic data such as gender, ethnicity, sociodemographic status, or educational background, which is an important limitation of this study. This makes it difficult to explore biases in the model, which may result in differing performance across demographic groups. Further work will need to be done to explore the algorithm's performance in a diverse range of groups to guard against differential performance. In addition, this limitation makes it hard for us to generalize these findings to a specific group outside of users of our app. However, the MAQ/MDQ are based on widely accepted symptoms that are likely to maintain their validity outside of this setting. Another limitation of this study was the relatively small sample size; future investigations could expand to larger data sets. In this vein, an important future step is to test the effectiveness of this algorithm in a setting with participants with well-labelled characteristics—an essential follow-on to the initial validation [29]. A future trial is planned to assess this in a group of older adults, an especially important group, since evaluating depression and anxiety are in the context of dementia risk factors and screening.

Conclusion

In summary, our results suggest that by using the PHQ-4, in line with the other measures collected in the Mindstep app, we can achieve accuracies similar to full-length PHQ-9 and GAD-7 questionnaires. This suggests that the app can be used to reliably screen for these conditions. Further work in populations with validated diagnoses whose demographics are known will further strengthen the evidence underlying these models.

Authors' Contributions

RPRZ and NK conducted the data analysis and devised this project. RPRZ and NK wrote the manuscript. MM, MJ, IP, HS, YY, and AAM reviewed the draft manuscript and provided editorial input.

Conflicts of Interest

All authors are paid employees of Mindset Technologies Ltd.

Multimedia Appendix 1
Supplementary information on the performance of the machine learning models in the training set.
[DOCX File , 14 KB - formative_v6i3e31209_app1.docx]

References

1. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. The Lancet 2020 Aug;396(10248):413-446. [doi: 10.1016/s0140-6736(20)30367-6]
2. Bennett S, Thomas AJ. Depression and dementia: cause, consequence or coincidence? Maturitas 2014 Oct;79(2):184-190. [doi: 10.1016/j.maturitas.2014.05.009] [Medline: 24931304]
3. Beats BC, Sahakian BJ, Levy R. Cognitive performance in tests sensitive to frontal lobe dysfunction in the elderly depressed. Psychol Med 1996 May;26(3):591-603. [doi: 10.1017/s0033291700035662] [Medline: 8733217]
4. Ballard C, Neill D, O'Brien J, McKeith I, Ince P, Perry R. Anxiety, depression and psychosis in vascular dementia: prevalence and associations. Journal of Affective Disorders 2000 Aug;59(2):97-106. [doi: 10.1016/s0165-0327(99)00057-9]
5. Eysenck MW, Derakshan N, Santos R, Calvo MG. Anxiety and cognitive performance: attentional control theory. Emotion 2007 May;7(2):336-353. [doi: 10.1037/1528-3542.7.2.336] [Medline: 17516812]
6. Elfgren C, Gustafson L, Vestberg S, Passant U. Subjective memory complaints, neuropsychological performance and psychiatric variables in memory clinic attendees: a 3-year follow-up study. Arch Gerontol Geriatr 2010;51(3):e110-e114. [doi: 10.1016/j.archger.2010.02.009] [Medline: 20211500]
7. Welzel FD, Stein J, Röhr S, Fuchs A, Pentzek M, Mösch E, et al. Prevalence of Anxiety Symptoms and Their Association With Loss Experience in a Large Cohort Sample of the Oldest-Old. Results of the AgeCoDe/AgeQualiDe Study. Front Psychiatry 2019;10:285 [FREE Full text] [doi: 10.3389/fpsyt.2019.00285] [Medline: 31139097]
8. Kroenke K, Spitzer R, Williams J. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
9. Yesavage JA, Brink T, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: A preliminary report. Journal of Psychiatric Research 1982 Jan;17(1):37-49. [doi: 10.1016/0022-3956(82)90033-4]
10. Phelan E, Williams B, Meeker K, Bonn K, Frederick J, Logerfo J, et al. A study of the diagnostic accuracy of the PHQ-9 in primary care elderly. BMC Fam Pract 2010 Sep 01;11:63 [FREE Full text] [doi: 10.1186/1471-2296-11-63] [Medline: 20807445]
11. Zhang H, Wang S, Wang L, Yi X, Jia X, Jia C. Comparison of the Geriatric Depression Scale-15 and the Patient Health Questionnaire-9 for screening depression in older adults. Geriatr Gerontol Int 2020 Feb;20(2):138-143. [doi: 10.1111/ggi.13840] [Medline: 31820572]
12. Wild B, Eckl A, Herzog W, Niehoff D, Lechner S, Maatouk I, et al. Assessing generalized anxiety disorder in elderly people using the GAD-7 and GAD-2 scales: results of a validation study. Am J Geriatr Psychiatry 2014 Oct;22(10):1029-1038. [doi: 10.1016/j.jagp.2013.01.076] [Medline: 23768681]
13. Vasiliadis H, Chudzinski V, Gontijo-Guerra S, Préville M. Screening instruments for a population of older adults: The 10-item Kessler Psychological Distress Scale (K10) and the 7-item Generalized Anxiety Disorder Scale (GAD-7). Psychiatry Res 2015 Jul 30;228(1):89-94. [doi: 10.1016/j.psychres.2015.04.019] [Medline: 25956759]
14. Plummer F, Manea L, Trepel D, McMillan D. Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. Gen Hosp Psychiatry 2016;39:24-31. [doi: 10.1016/j.genhosppsych.2015.11.005] [Medline: 26719105]
15. Hoerger M. Participant dropout as a function of survey length in internet-mediated university studies: implications for study design and voluntary participation in psychological research. Cyberpsychol Behav Soc Netw 2010 Dec;13(6):697-700 [FREE Full text] [doi: 10.1089/cyber.2009.0445] [Medline: 21142995]
16. Treadway M. The Neurobiology of Motivational Deficits in Depression--An Update on Candidate Pathomechanisms. Curr Top Behav Neurosci 2016;27:337-355. [doi: 10.1007/7854_2015_400] [Medline: 26475160]
17. Keller AS, Leikauf JE, Holt-Gosselin B, Staveland BR, Williams LM. Paying attention to attention in depression. Transl Psychiatry 2019 Nov 07;9(1):279 [FREE Full text] [doi: 10.1038/s41398-019-0616-1] [Medline: 31699968]
18. Kroenke K, Spitzer RL, Williams JBW, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. Psychosomatics 2009;50(6):613-621. [doi: 10.1176/appi.psy.50.6.613] [Medline: 19996233]

19. Bhana A, Rathod SD, Selohilwe O, Kathree T, Petersen I. The validity of the Patient Health Questionnaire for screening depression in chronic care patients in primary health care in South Africa. BMC Psychiatry 2015 May 23;15:118 [FREE Full text] [doi: 10.1186/s12888-015-0503-0] [Medline: 26001915]

20. Levis B, Sun Y, He C, Wu Y, Krishnan A, Bhandari PM, Depression Screening Data (DEPRESSD) PHQ Collaboration, et al. Accuracy of the PHQ-2 Alone and in Combination With the PHQ-9 for Screening to Detect Major Depression: Systematic Review and Meta-analysis. JAMA 2020 Jun 09;323(22):2290-2300 [FREE Full text] [doi: 10.1001/jama.2020.6504] [Medline: 32515813]

21. Ewing JA. Detecting Alcoholism. JAMA 1984 Oct 12;252(14):1905. [doi: 10.1001/jama.1984.03350140051025]

22. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. CMAJ 2012 Feb 21;184(3):E191-E196 [FREE Full text] [doi: 10.1503/cmaj.110829] [Medline: 22184363]

23. Arik S, Pfister T. TabNet: attentive interpretable tabular learning. arXiv.org. URL: http://arxiv.org/abs/1908.07442 [accessed 2021-06-11]

24. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From Local Explanations to Global Understanding with Explainable AI for Trees. Nat Mach Intell 2020 Jan;2(1):56-67 [FREE Full text] [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]

25. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep;16(3):297-334. [doi: 10.1007/BF02310555]

26. Bavelloni A, Piazzi M, Raffini M, Faenza I, Blalock WL. Prohibitin 2: At a communications crossroads. IUBMB Life 2015 Apr;67(4):239-254 [FREE Full text] [doi: 10.1002/iub.1366] [Medline: 25904163]

27. Trinh NT, Youn SJ, Sousa J, Regan S, Bedoya CA, Chang TE, et al. Using electronic medical records to determine the diagnosis of clinical depression. Int J Med Inform 2011 Jul;80(7):533-540 [FREE Full text] [doi: 10.1016/j.ijmedinf.2011.03.014] [Medline: 21514880]

28. Wang S, Pathak J, Zhang Y. Using Electronic Health Records and Machine Learning to Predict Postpartum Depression. Stud Health Technol Inform 2019 Aug 21;264:888-892. [doi: 10.3233/SHTI190351] [Medline: 31438052]

29. Rifkin-Zybutz R, Selim H, Johal M, Kuleindiren N, Palmon I, Lin A, et al. Preliminary validation study of the Mindset4Dementia application: assessing remote collection of dementia risk factors and cognitive performance. BMJ Innov 2021 Sep 17;7(4):626-631. [doi: 10.1136/bmjinnov-2021-000780]

Abbreviations

AUC: area under the curve
CAGE: Cut, Annoyed, Guilty, and Eye Questionnaire
GAD-7: 7-item Generalized Anxiety Disorder Scale
MAE: mean absolute error
MDQ: Mindset Depression Question
MAQ: Mindset Anxiety Question
PHQ-9: 9-item Patient Health Questionnaire for depression
SCID: Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders
SHAP: Shapley Additive Explanations

XSL•FO

RenderX

Original Paper

Recommendations for Researchers on Synchronous, Online, Nominal Group Sessions in Times of COVID-19: Fishbone Analysis

Lotte Timmermans[1], MSc; Ine Huybrechts[2], MSc; Peter Decat[3], MSc, PhD; Veerle Foulon[4], MSc, PhD; Ann Van Hecke[5], MSc, PhD; Mieke Vermandere[1], MD, PhD; Birgitte Schoenmakers[1], MD, PhD

[1]Academic Centre of General Practice, KU Leuven, Leuven, Belgium

[2]Department of Family Medicine and Population Health, University of Antwerp, Antwerp, Belgium

[3]General Practice and Primary Health Care, Ghent University, Ghent, Belgium

[4]Clinical Pharmacology and Pharmacotherapy, KU Leuven, Leuven, Belgium

[5]University Centre for Nursing and Midwifery, Ghent University, Ghent, Belgium

Corresponding Author:
Lotte Timmermans, MSc
Academic Centre of General Practice
KU Leuven
Kapucijnenvoer 7
Box 7001
Leuven, 3000
Belgium
Phone: 32 16 37 45 98
Email: lotte.timmermans@kuleuven.be

Abstract

Background: In times of COVID-19, we are challenged to experiment with alternative platforms or software to connect people. In particular, the struggle that arose in health research was how to interact with patients and care professionals. The latter is additionally faced with an extreme workload to fight the pandemic crisis. Creative strategies have been developed to continue research among patients and care professionals to improve quality of care. This paper addresses the issue of synchronous, online, nominal group sessions, a common consensus method used for group brainstorming.

Objective: The purpose of this study was to share our experiences with performing online, nominal group sessions using the video conference software Microsoft Teams. In addition, we aimed to create a practical guide with recommendations for researchers.

Methods: We critically analyzed the procedures for the online nominal group technique, according to the Fishbone methodology.

Results: Performing synchronous, online, nominal group sessions is challenging but offers opportunities. Although interaction with and among the attendees complicates the process, the major advantage of online sessions is their accessibility and comfort because of reduced barriers to participation (eg, lower time investment). The role of the moderators is of major importance, and good preparation beforehand is required. Recommendations for future online, nominal research were formulated.

Conclusions: Online, nominal group sessions seem to be a promising alternative for the real-life commonly used technique. Especially during the COVID-19 pandemic, the benefits must be highlighted. More expertise is needed to further refine the practical guide for using digital software in research and to achieve optimal performance.

KEYWORDS

COVID-19; fishbone diagram; nominal group technique; video conferencing; primary health care; qualitative research

Introduction

The COVID-19 outbreak has posed multiple challenges. Not only has the pandemic caused major loss of human life globally, the impact on mental and psychosocial health should not be underestimated [1-4]. Apart from the major impact on human health, it also has an impact on health research. More specifically, the limited accessibility of research sources and

XSL•FO
RenderX

participants obliges researchers to seek for alternatives to pursue research projects. Due to strict regulations to control transmission of the virus, researchers have to rely on creative strategies to involve the research population [5]. In-depth interviews and focus groups are conducted by using video conferencing software and platforms, but more complex group-based participatory methods have not yet been fully explored online [6-8]. The nominal group technique is one of these unexplored methodologies in research. This technique is commonly used to solve problems, determine priorities, or generate ideas [9]. The main challenge is cooperation in a strict and formal procedure [10]. Guidance is provided to a limited extent for face-to-face sessions, but the online setting remains unclear [9-11]. There are 2 different strategies for organizing online sessions: asynchronous sessions using an online platform to communicate in a nonsimultaneous way or synchronous sessions using video conference software in which participants join simultaneously. Since the nominal group methodology has not yet been extended fully online, very few researchers have utilized it [12,13]. Guidelines for recruiting participants, giving informed consent, and doing qualitative research with individuals are all based on the assumption of in-person interactions [14]. Therefore, there is a need to investigate in-depth the methodological issues of an online setup. In this paper, we describe the challenges and opportunities of the online form based on our experiences with 3 synchronous, virtual, nominal group sessions. In addition, participants' experiences were explored. Finally, a practical guide with recommendations for researchers has been created.

The online sessions were part of a larger study by the Primary Care Academy (PCA) that aims to implement self-management support into primary care practice. Self-management is "the individual's ability to manage the symptoms, treatment, physical and psychosocial consequences and lifestyle changes inherent in living with a chronic condition" [15]. Self-management of a chronic disease is a process in which health care professionals play a crucial supportive role [16-18]. The COVID-19 pandemic has highlighted the importance of self-management since the lockdown of many services resulted in more responsibility regarding handling of chronic diseases [19]. Online nominal group sessions were organized to brainstorm how health care professionals can support patients in this self-management process.

Methods

We critically analyzed different aspects of synchronous, online, nominal group sessions, according to the Fishbone methodology. The nominal group sessions aimed to formulate specific actions to support patient's self-management and were part of research to reinforce Flemish primary care.

Data Collection

Data were collected during 3 online, nominal group sessions with health care professionals and experts by experience (ie, patients with knowledge based on personal experiences). Two qualified moderators conducted all the sessions, by use of the video conference software Microsoft Teams. With the participants' consent, all 3 sessions were audio- and video-recorded and transcribed verbatim afterwards for qualitative analysis. At the end of each session, participants were offered the opportunity to share their experiences verbally, by use of the chat box, or by email afterwards.

Online, Nominal Group Sessions

Similar to the classical nominal group technique, the online sessions were structured in 4 key stages: silent generation, round robin, clarification, and ranking [9]. A PowerPoint presentation was used to support this structure and the accompanying storyline. In advance, this slideshow was uploaded in Microsoft Teams instead of sharing the presentation, with the advantage of seeing the participants on the call. The protocol was developed specifically for online use and based upon comprehensive debate within the research group.

Sampling and Recruitment

Maximum variation purposive sampling was used to recruit the nominal group participants by email or through health care organizations and practices. A registration link was embedded in a flyer. In addition, a call for participants was announced on the internet (website of the PCA). Participants included health care professionals and experts by experience from around Flanders. The aim was to have a minimum of 5 and a maximum of 10 participants per session [20].

Procedure

Each session started with a short introduction, including explaining ground rules (which had also been communicated in advance by email) and a roundtable introduction by the attendees. Afterwards, information was provided on the topic of the session (self-management support) and the brainstorming procedure. These initial stages took 20 minutes and were followed by the active nominal group brainstorming. It started with the presentation of a specific question, related to self-management support, that gave rise to the formulation of ideas during the silent generation phase. This idea generation question (ie, "What concrete action/interventions can be designed to help patients fit chronic disease(s) into their lives in the most optimal way?") was the impetus for brainstorming. Participants were asked in advance to have their pen and paper ready while brainstorming. To keep everyone's attention, no more than 15 minutes were provided for generating ideas on self-management support. Additionally, the round robin stage in which ideas were exchanged filled in the next 20 minutes. Both moderators took notes during this stage. The formulated ideas were then rephrased by the moderators to make sure everyone gained a full understanding. Then, during a 15-minute clarification phase, there was time to clarify the ideas upon request of the participants. Finally, the nominal group sessions ended with a ranking stage of 5 minutes to 10 minutes, depending on the number of ideas formulated. More specifically, a ranking system created with the online service "Poll Everywhere" (San Francisco, CA) was included in the slideshow to prioritize the ideas generated according to the participants' preferences. A short break of 10 minutes after the generation of ideas gave the moderators the opportunity to process them in the ranking system. The second and third sessions had an additional phase just before the polling in which the ideas

previously formulated were quickly reviewed and included in the ranking system. Each session was planned to last no longer than 1.5 hours.

Analysis of the Content Related to Self-Management Support

Processing the output of the brainstorming sessions (ie, formulated ideas on self-management support) was performed analogous to the regular methods of the nominal group technique. In summary, the generated ideas were (after performing 3 brainstorming sessions) organized in themes by qualitative coding of the transcripts and processed in a survey. Participants were asked to identify their favorite idea and to rank the ideas based on their preferences by defined theme. Afterwards, ideas were grouped by the main researcher (LT) following a self-management support model into 5 categories according to the type of primary care action (ie, supporting, involving, listening to, coordinating, or questioning patients). Ideas that were not chosen as favorites were excluded. Subsequently, the table was reviewed by the research team during multiple rounds to reach a consensus on the categorization. Finally, ideas were further refined according to the survey results and processed for use in a primary care practice intervention related to self-management support. The results of the analysis related to the formulated ideas of the online sessions are beyond the scope of this research paper and will be reported elsewhere.

Ethical Approval

Ethical approval for the brainstorming sessions was obtained from The Ethics Committee Research UZ/KU Leuven (S63890). An informed consent document was sent to the participants in advance by email. It explicitly stated that conversations were video- and audio-recorded during the online sessions. In addition, participants were informed that the content remained internal for scientific research purposes.

The return of a signed copy was required to participate in the online sessions. After confirmation of participation and signing the informed consent form, attendees obtained an invitation link to a Microsoft Teams live event. Access to the recorded sessions (audio, video, and shared slideshow) was limited to the research group. The entire study was conducted in accordance with the Helsinki Declaration.

Outcome Measures

We defined 6 different process outcomes to characterize the online nominal group sessions: guidance, engagement, interaction, timing, technology, and participants' experiences. These categories were initially described by the main researcher based on literature analysis and further refined by the research team.

Analysis of the Online, Nominal Group Procedure

Audio and video recordings, together with written data (ie, transcripts, Teams chat, and email data) from the nominal group procedure, were analyzed for the purpose of this study. A deductive "top-down" approach was applied to collect meaningful data from these sources, by exploring the predefined outcomes. More specifically, a Fishbone diagram was used to

explore all possible challenges researchers and participants faced when performing or attending the online, nominal group sessions, as this methodology helps team members visualize the potential problem sources. The analysis starts by defining a central problem, which is then deeply examined by formulating several causes, organized in different categories [21]. A literature search revealed that an online, nominal group technique has rarely been performed (and reported); only very few research articles explored this online methodology. This resulted in the definition of our central problem: low uptake of synchronous, online, nominal group sessions in research practice. The final Fishbone diagram incorporated in the study was reviewed by the entire research team to manage bias and ensure the validy of our findings.

Trustworthiness

To ensure rigor, we applied different strategies to increase trustworthiness of the data. First, maximum variation sampling was applied to recruit the participants for the nominal group sessions. Participants of different ages, sexes, regions, educational levels, and employment were chosen. Second, the 3 nominal group sessions were fully recorded and transcribed verbatim. In addition, field notes were taken. Third, dependability of the findings was established by providing a detailed description of the online, nominal group procedure. Finally, the analysis was checked by qualitative experts in the field.

Research Team

LT and IH performed the online nominal group sessions and collected the data. LT analyzed the data and designed the Fishbone diagram, in close collaboration with supervisor BS. The entire research team (IH, PD, VF, AVH, MV, and BS) reviewed the final diagram that was incorporated in this manuscript. This group consists of experienced members in qualitative health research. Before project initiation, the moderators received additional intensive training on the principles and methods in qualitative research to assure a certain level of standardization. Both moderators had previous expertise in quantitative and qualitative data collection in a group setting.

Results

After performing 3 synchronous, online, nominal group sessions in which individuals brainstormed on specific actions to support patients' self-management, a Fishbone analysis was undertaken. A total of 24 persons participated in the online sessions and therefore contributed to the data collection.

Challenges for Online, Nominal Group Sessions

Various causes were identified that positively or negatively challenged the online performance. These causes were clustered under the predefined outcome measures: guidance, engagement, interaction, timing, technology, and participants' experiences. Each outcome measure was thoroughly examined by audio and video analysis, supplemented with input from the chat and email, to identify the potential causes of the low uptake of online, nominal group sessions in research practice. Figure 1 represents the final Fishbone diagram.

Figure 1. Fishbone diagram representing identified causes challenging the uptake of synchronous, online, nominal group sessions. Template retrieved from an online collaborative whiteboard [22].
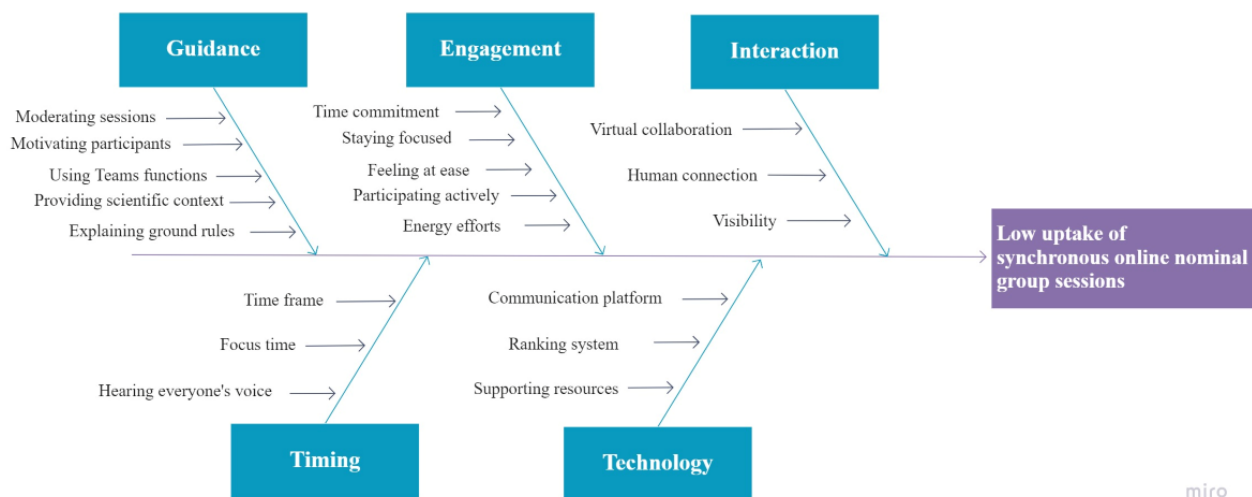
Guidance

Moderating the online sessions challenged the research team in various ways. The moderators' role was of major importance since they had to guide the discussion in an online setting. The main barrier was managing participants during the brainstorming while not being able to interact directly, neither verbally nor nonverbally. Participants were encouraged to provide input by elevating a virtual hand using the hand icon of Microsoft Teams. A moderator then appointed the person to speak at an appropriate time during the session. The importance of guiding participants was especially expressed during the round robin and clarification stages. To ensure everyone could provide their input, participants were indicated one by one to express their ideas on the topic of self-management. Examples of formulated ideas included the following: introducing a buddy system among peers, organizing interactive sessions on chronic disease coping strategies, integrating communication aids and strategies in primary care practice. This structured approach while exchanging the ideas seemed to have positively contributed to an equal involvement of all participants, even those who had told us beforehand they were not very articulate. Questions to each other when sharing ideas could often not be answered immediately. In these cases, the use of a chat box could act as an intermediate communication medium. Although instructed, this medium was rarely used by the participants.

During the entire nominal group session, attention was paid to expressing the scientific context and methodology in simple terms—for example, the moderators did not use the word self-management itself when guiding conversations and asking the idea generation question. This resulted in an overall understanding of the background information, setup, and aim of the nominal group technique. Based on the participants' feedback we received, we can conclude that the absence of real-life interaction emphasized the importance of strictly guided sessions. Also, clear instructions before and at the start were necessary given the simultaneous use of the chat box, hand icon, and PowerPoint presentation by the moderator, while interacting with the audience.

Engagement

During the sessions, engaging participants was challenging because of the online setup. Some actions by the moderators were observed as a remarkable help, since they triggered the audience to contribute to the conversation. First, the sessions began with a roundtable introduction, encouraging participants to get to know each other in an accessible way. This introduction revealed a variety of backgrounds: project officers of care organizations, general practitioners, nurses, patients with chronic conditions and informal caregivers, representatives from patient associations, lecturers, and researchers in primary care. Second, the use of informative PowerPoint slides during the idea generation phase seemed a welcome addition, since all participants indicated they had a full understanding of the brainstorming procedure after reading a concise information sheet on the general topic and the idea generation question. Third, to give participants time to think and feel at ease, they were asked to switch off the camera during the silent idea generation phase. Fourth, regarding the engagement while exchanging ideas, the use of a round-robin questioning format allowed each participant to provide input. Attendees were motivated to actively participate by emphasizing that all input was valuable. However, we noticed the online setup demanded extra effort from the participants. This was mainly observed at the end of the sessions, when there was reduced active contribution to the brainstorming. This could be attributed to the long period of attention on the computer screen.

Interaction

The online design challenged the interactive process of brainstorming since participants could not rely on human interaction and the use of sticky notes, both key elements in real-life nominal group sessions. Nevertheless, some elements did contribute to a relatively vibrant process. Slides with strong visuals, including images to supplement content, engaged participants online and resulted in virtual collaboration. Participants also used pen and paper to write down their thoughts. In addition, the hand icon was used to interact with the research team. Unfortunately, active interaction or conversation among participants was not enhanced by the online setup. The only stage in which participants were able to

communicate with each other in a coordinated way was the clarification stage.

A remarkable aspect of the online setup was how the camera use influenced the level of interaction. A sense of interconnectedness was created when participants were asked to switch on their webcams, determined by the observation that attention increased and the conversation became more lively. Noteworthy is that when sharing the PowerPoint presentation, this connected feeling diminished as the screen was taken up by the shared slides instead of the video streams. To overcome this barrier, screensharing was interrupted while actively brainstorming during the round robin and clarification stages.

Timing

Overall, the timing of the virtual sessions turned out well, although moderators were challenged to stick to the schedule in the foreseen period. To maintain the timetable, the entire nominal group session was precisely structured, and every part was delineated according to a strict time frame. The timing was tuned to the PowerPoint slides. For the moderators, staying within the time frame was rather challenging since they had to find a balance between providing enough time versus sticking to the strict time frame. To make sure sessions were not too strenuous, a break of 10 minutes was incorporated in between and appeared to be sufficient. However, the time frame was very tight, resulting in a rushed feeling among the moderators and reduced room for unforeseen circumstances (eg, delayed start). We also noticed that there was not enough time to prioritize the ideas with the external ranking system. To resolve this, we provided the opportunity to rank immediately afterwards up to a few hours after each session. However, this function was rarely used.

Technology

The online nominal group design challenged the participants since they had to have access to reliable technology to successfully engage within the sessions. In addition, necessary skills for using the various instruments (ie, Microsoft Teams, Poll Everywhere) were expected. Likewise, the moderators had to be able to manage the interaction with the audience, processing of ideas (ie, submitting in a ranking system), and navigation through the slideshow. This emphasized the need for guidance from 2 moderators. Unfortunately, due to the large number of participants (between 7 and 10 in each session), not all the faces were visible at the video bar while presenting the slides. This was a disadvantage because visual interaction appeared to contribute to participants' engagement.

Participants' Experiences

I would like to compliment you (ie, the main moderator/researcher) on the way you took the lead in the session. You gave enough guidance so that it was pleasant to follow at a smooth pace. That is not always evident online. [Participant]

Based on voluntary feedback in the chat, by email, or orally, participants' experiences regarding the online sessions were positive overall. Furthermore, participants explicitly indicated they wished to cooperate further on the research project. The only negative feedback was received regarding the large number of ideas that had to be prioritized using the ranking system (Poll Everywhere). Participants verbally indicated that they did not have enough time to rank the ideas in the foreseen time frame. This resulted in confusion for some participants and failure to reach a full consensus at the end of each session, since consensus can only be reached when everyone gives their input. It is possible that participants preferred to have more time in a relaxed atmosphere to rank ideas. This resulted in the development of an additional short survey in Qualtrics (Provo, UT) after completion of the 3 sessions, in which all participants had the opportunity to cast their vote on the entire group of ideas (total of 3 sessions).

Discussion

Practical Guide With Recommendations

Guidance for performing online nominal group studies is missing. However, it has been stated that results of the same quality can be achieved with an online approach [23]. In addition, studies show that there is no specific need to perform real-life sessions [13]. We used a Fishbone diagram to identify all possible challenges that researchers face when performing online, nominal group sessions. Based on our findings and experiences with the online setting, we developed a practical guide with recommendations for researchers interested in this type of work. The recommendations were pooled into the main stages of the nominal group: silent generation, round robin, clarification, and ranking (Figure 2). In addition, some general recommendations were formulated.

Figure 2. Recommendations for planning and implementing synchronous, online, nominal group sessions.

Recommendations for online nominal group sessions

Silent generation
1. Ask to have pen and paper available
2. Display initial question on screen
3. Make sure everyone understands the concept
4. Keep an eye on the chat box
5. Ask to turn off the camera
6. Set up a short time frame

Round robin
1. Provide enough time
2. Stop sharing slideshow and focus on interaction (camera on)
3. Discourage the use of the chat box
4. Act as a neutral moderator
5. Let each person pass his or her ideas to the group
6. Coordinate by indicating participants

Clarification
1. Make sure everyone understands the ideas
2. Rephrase the ideas
3. Stimulate the use of the hand icon
4. Check chat box regularly
5. Determine the end point in advance

Ranking
1. Present a clear overview of the generated ideas
2. Include a user-friendly ranking system
3. Give clear and understandable instructions
4. Finish by reporting some results

XSL·FO
RenderX

Silent Generation

At the beginning of the session, the moderators should clearly explain the concept. During the brainstorming, pen and paper are indispensable. Taking notes helps the participants to explore ideas. Instead of classical notes, an online platform or whiteboard can help participants brainstorm but can be more complicated. Displaying the initial question on the screen during the generation process adds value. We recommend keeping this phase short, to avoid losing someone's attention. Moderators should encourage participants to use the chat box when personal issues occur. Oral interventions are not recommended. This results in disruption of the individual reflection. An exception should be made for communications that are relevant to the whole group. In addition, switching off the cameras also contributes to thinking in serenity. However, disabling cameras can be discussed as a negative element, as participant's behavior can no longer be interpreted. In addition, seeing each other visually can result in increased commitment to the idea generation.

Round Robin

To engage participants during the round robin, we recommend stopping screen sharing. Switching on the video camera leads to a more interactive conversation. To further increase the level of interaction, try to minimize the use of the chat box while exchanging ideas. If still used during this stage, motivate the chat box user to ask the question out loud. We suggest providing enough time to exchange the ideas generated so that everyone can have their say and the moderators have time to take notes. Furthermore, it is essential that moderators act as neutral as possible to avoid judgment and criticism. They should monitor the participants' engagement. Coordination is crucial for every person to pass their ideas to the group.

Clarification

Collaboration and interaction are highly affected in the online setup since there is less nonverbal communication. This challenges participants to speak more out loud during the clarification round. Unfortunately, not everyone feels comfortable doing so. The use of the chat box as an intermediate communication platform can counteract this but was rarely used. Moderators should make sure everyone understands the ideas formulated. It helps to reformulate ideas in different ways and to check that those present understand them. Make sure the use of the Microsoft Teams hand icon is encouraged to avoid a chaotic conversation. Analogous to the regular methods of the nominal group technique, we advise determining the end point of this research phase in advance to prevent clarification from becoming an ongoing process.

Ranking

Some participants are overwhelmed by a multitude of information. Providing a clear overview of the generated ideas at the beginning of the ranking process is helpful. This procedure should be user-friendly, and instructions should be incorporated into the accompanying slideshow. The time it took to rank the ideas generated seemed to be person-dependent. Therefore, it is valuable to familiarize oneself with the audience while preparing the online sessions. To finish the brainstorming in a pleasant way, let the participants take a sneak peek at the results. This can be done by reporting the most favorite idea so far (ie, ranked as number 1).

General Recommendations

Organizing synchronous, online, nominal group sessions requires thorough preparation. Guidance by at least two moderators is necessary. Due to the online complexity, moderators can benefit from special training. Not only are skills required from the research team but participants also need to develop online competencies. Unfortunately, not everyone is able to deal with online platforms. In Belgium, more than 15% of the adult population has a low level of health literacy, and in addition, 10% of households do not have an internet connection [24,25]. This must be considered when setting up online research with lay participants. Communication with the participants is essential to operate efficiently. More specifically, moderators should provide clear instructions before, at the start of, and during sessions. Furthermore, it is important to make a detailed schedule and monitor time. To engage participants as much as possible, more emphasis should be put on enhancing human interaction. The format of the nominal group technique is well adapted to empower attendees [13,26]. However, it should be mentioned that due to a strict pattern and guidelines, the nominal group technique never allows a lot of interaction. Not being able to ask each other questions immediately, due to the strict design, might also challenge participants' patience. Starting with a roundtable introduction, switching on the video cameras, and using interactive slides with visual components seem to make attendees feel comfortable. In addition, the possibility can be offered to contact the research team afterwards. Researchers should aim for a meeting moment that works best for the target population, and the importance of attending at the scheduled time needs to be emphasized.

Computer screens have been proven to negatively influence the activity of the human brain [27]. Therefore, the duration of the meeting should be limited. By contrast, online sessions are less subject to distractions, which means less time is lost and researchers can better stick to the schedule. Unfortunately, the strict timing can result in a rushed feeling. In our opinion, this cannot be compensated by allowing more time for the brainstorming, as we noticed focus decreased the longer the session lasted. A second break could be a possible solution.

The main advantage of performing online sessions seems to be the flexibility and comfort for both researchers and attendees. The online setup increased the accessibility for participants who otherwise might have experienced barriers to participation, such as feelings of discomfort in a group setting, transportation issues, or time [13]. Especially for health care professionals, who were in the middle of the battle during COVID-19 pandemic, it was a great opportunity to actively participate in research. The flexibility might explain the large turnout and few cancellations. This is in striking contrast to real-life sessions in which organizational issues seem the main limitation [10].

Strengths and Limitations

Some limitations should be mentioned. First, performing only 3 online sessions limited the generalizability of the results.

However, after performing 3 sessions, we reached a point of saturation regarding both the findings on the topic of self-management support and on the methodological challenges and opportunities. Moreover, we did not compare our findings with real-life nominal group sessions. This paper only elaborates on the synchronous, online technique. Based on the performance of 3 sessions, we were able to gain sufficient input to analyze challenges. Knowledge of the real-life setup was gained through literature analysis. Furthermore, participants' experiences were not deeply surveyed or explored. However, we achieved a notion of participants' satisfaction by voluntary feedback in the chat, by mail, or orally. In addition, the research team reflected critically about the objectivity of the observations during multiple review rounds. Finally, the use of sophisticated online brainstorming tools was not included in our brainstorming procedure. These platforms can increase engagement during brainstorming. However, we deliberately chose simple operating systems (ie, Microsoft Teams, Poll Everywhere), because not everyone is able to work with such tools. By choosing low-threshold systems, a wide target group could participate in the online, nominal group sessions.

Practical Implications

Online, nominal group sessions seem to be a promising alternative to the commonly used real-life technique. This paper provides researchers with recommendations for conducting online sessions, considering the various challenges. In our experience, the online format is highly recommended when looking for research procedures that are very accessible and consume little time. Especially during the COVID-19 pandemic, the benefits must be highlighted.

Future Research

Future research should focus on refining the online, nominal group technique. More expertise is needed to optimize the practical guide and to achieve optimal performance. Comparative studies between the real-life and virtual setups are required to make statements about the efficiency. Researchers should consider participants' experiences in the design of future online sessions.

Conclusion

Performing synchronous, online, nominal group sessions is challenging but offers opportunities. One should be aware of the differences between real-life and online sessions. Good preparation is needed to overcome the barriers of the online technique. Our practical guide for researchers offers recommendations to facilitate the process. The role of the moderators is of major importance during brainstorming. Further research should refine the online procedure and make it more accessible for both researchers and the research population.

Conflicts of Interest

None declared.

References

1. Sethi BA, Sethi A, Ali S, Aamir HS. Impact of Coronavirus disease (COVID-19) pandemic on health professionals. Pak J Med Sci 2020 May 18;36(COVID19-S4):1. [doi: 10.12669/pjms.36.covid19-s4.2779]

2. Xiong J, Lipsitz O, Nasri F, Lui LM, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: A systematic review. J Affect Disord 2020 Dec 01;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]

3. Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): A review. Int J Surg 2020 Jun;78:185-193 [FREE Full text] [doi: 10.1016/j.ijsu.2020.04.018] [Medline: 32305533]

4. Tan E, Song J, Deane A, Plummer M. Global impact of coronavirus disease 2019 infection requiring admission to the ICU: a systematic review and meta-analysis. Chest 2021 Feb;159(2):524-536 [FREE Full text] [doi: 10.1016/j.chest.2020.10.014] [Medline: 33069725]

5. Education responses to COVID-19: Embracing digital learning and online collaboration. OEDC. 2020 Mar 23. URL: https://www.oecd.org/coronavirus/policy-responses/education-responses-to-covid-19-embracing-digital-learning-and-online-collaboration-d75eb0e8/ [accessed 2022-03-14]

XSL•FO

RenderX

6. Dos Santos Marques IC, Theiss LM, Johnson CY, McLin E, Ruf BA, Vickers SM, et al. Implementation of virtual focus groups for qualitative data collection in a global pandemic. Am J Surg 2021 May;221(5):918-922 [FREE Full text] [doi: 10.1016/j.amjsurg.2020.10.009] [Medline: 33070983]

7. Halliday M, Mill D, Johnson J, Lee K. Let's talk virtual! Online focus group facilitation for the modern researcher. Res Social Adm Pharm 2021 Dec;17(12):2145-2150. [doi: 10.1016/j.sapharm.2021.02.003] [Medline: 33653681]

8. Jones RE, Abdelfattah KR. Virtual interviews in the era of COVID-19: a primer for applicants. J Surg Educ 2020 Jul;77(4):733-734 [FREE Full text] [doi: 10.1016/j.jsurg.2020.03.020] [Medline: 32278546]

9. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. Int J Clin Pharm 2016 Jun 5;38(3):655-662 [FREE Full text] [doi: 10.1007/s11096-016-0257-x] [Medline: 26846316]

10. Waggoner J, Carline JD, Durning SJ. Is there a consensus on consensus methodology? Descriptions and recommendations for future consensus research. Academic Medicine 2016;91(5):663-668. [doi: 10.1097/acm.0000000000001092]

11. Allen J, Dyas J, Jones M. Building consensus in health care: a guide to using the nominal group technique. Br J Community Nurs 2004 Mar;9(3):110-114. [doi: 10.12968/bjcn.2004.9.3.12432] [Medline: 15028996]

12. Tseng K, Lou S, Diez C, Yang H. Using online nominal group technique to implement knowledge transfer. J Eng Educ 2006;95(4):335-345. [doi: 10.1002/j.2168-9830.2006.tb00908.x]

13. Kulczycki A, Shewchuk RM. Using Internet-based nominal group technique meetings to identify provider strategies for increasing diaphragm use. J Fam Plann Reprod Health Care 2008 Oct 01;34(4):227-231 [FREE Full text] [doi: 10.1783/147118908786000550] [Medline: 18942235]

14. Dos Santos Marques IC, Herbey II, Theiss LM, Hollis RH, Knight SJ, Davis TC, et al. Understanding the surgical experience for African-Americans and Caucasians with enhanced recovery. Journal of Surgical Research 2020 Jun;250:12-22. [doi: 10.1016/j.jss.2019.12.034]

15. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Education and Counseling 2002 Oct;48(2):177-187. [doi: 10.1016/S0738-3991(02)00032-0]

16. Alvarez C, Greene J, Hibbard J, Overton V. The role of primary care providers in patient activation and engagement in self-management: a cross-sectional analysis. BMC Health Serv Res 2016 Mar 11;16:85 [FREE Full text] [doi: 10.1186/s12913-016-1328-3] [Medline: 26969293]

17. Lorig KR, Sobel DS, Ritter PL, Laurent D, Hobbs M. Effect of a self-management program on patients with chronic disease. Eff Clin Pract 2001;4(6):256-262. [Medline: 11769298]

18. Samuelson M, Tedeschi P, Aarendonk D, de la Cuesta C, Groenewegen P. Improving interprofessional collaboration in primary care: position paper of the European Forum for Primary Care. Qual Prim Care 2012;20(4):303-312. [Medline: 23113915]

19. Danhieux K, Buffel V, Pairon A, Benkheil A, Remmen R, Wouters E, et al. The impact of COVID-19 on chronic care according to providers: a qualitative study among primary care practices in Belgium. BMC Fam Pract 2020 Dec 05;21(1):255 [FREE Full text] [doi: 10.1186/s12875-020-01326-3] [Medline: 33278877]

20. Manera K, Hanson CS, Gutman T, Tong A. Consensus Methods: Nominal Group Technique. In: Liamputtong P, editor. Handbook of Research Methods in Health Social Sciences. Singapore: Springer Nature; 2019:737-750.

21. Phillips J, Simmonds L. Using fishbone analysis to investigate problems. Nurs Times 2013;109(15):18-20. [Medline: 23696993]

22. The visual collaboration platform for every team. Miro. URL: https://miro.com/ [accessed 2022-03-24]

23. Lago PP, Beruvides MG, Jian J, Canto AM, Sandoval A, Taraban R. Structuring group decision making in a web-based environment by using the nominal group technique. Computers & Industrial Engineering 2007 Mar;52(2):277-295. [doi: 10.1016/j.cie.2006.11.003]

24. Cijfers over geletterdheid. Vlaanderen. URL: https://onderwijs.vlaanderen.be/nl/cijfers-over-geletterdheid [accessed 2022-03-14]

25. Digitaal isolement: bijna een kwart van alleenstaanden heeft thuis geen internet. Statbel. 2020 Apr 30. URL: https://statbel.fgov.be/nl/nieuws/digitaal-isolement-bijna-een-kwart-van-alleenstaanden-heeft-thuis-geen-internet [accessed 2022-03-14]

26. Leslie M, Khayatzadeh-Mahani A, MacKean G. Recruitment of caregivers into health services research: lessons from a user-centred design study. Res Involv Engagem 2019;5:17 [FREE Full text] [doi: 10.1186/s40900-019-0150-6] [Medline: 31139432]

27. Small GW, Lee J, Kaufman A, Jalil J, Siddarth P, Gaddipati H, et al. Brain health consequences of digital technology use. Dialogues Clin Neurosci 2020 Jun;22(2):179-187 [FREE Full text] [doi: 10.31887/DCNS.2020.22.2/gsmall] [Medline: 32699518]

Abbreviations

PCA: Primary Care Academy

Feasibility and Acceptability of Ecological Momentary Assessment With Young Adults Who Are Currently or Were Formerly Homeless: Mixed Methods Study

Sara Semborski[1*], MSW; Benjamin Henwood[1*], MSW, PhD; Brian Redline[2], BA; Eldin Dzubur[1*], PhD; Tyler Mason[3], PhD; Stephen Intille[4,5], PhD

[1]Suzanne Dworak-Peck School of Social Work, University of Southern California, Los Angeles, CA, United States

[2]School of Medicine, Oregon Health and Science University, Portland, OR, United States

[3]Department of Population and Public Health Science, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

[4]Khoury College of Computer Sciences, Northeastern University, Boston, MA, United States

[5]Bouvé College of Health Sciences, Northeastern University, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Sara Semborski, MSW
Suzanne Dworak-Peck School of Social Work
University of Southern California
669 W 34th Street
Los Angeles, CA, 90089
United States
Phone: 1 (213) 740 2711
Fax: 1 (213) 821 2088
Email: semborsk@usc.edu

Abstract

Background: Ecological momentary assessment (EMA) has been used with young people experiencing homelessness to gather information on contexts associated with homelessness and risk behavior in real time and has proven feasible in this population. However, the extent to which EMA may affect the attitudes or behaviors of young adults who are currently or were formerly homeless and are residing in supportive housing has not been well investigated.

Objective: This study aims to describe the feedback regarding EMA study participation from young adults who are currently or were formerly homeless and examine the reactivity to EMA participation and compliance.

Methods: This mixed methods study used cross-sectional data collected before and after EMA, intensive longitudinal data from a 7-day EMA prompting period, and focus groups of young adults who are currently or were formerly homeless in Los Angeles, California, between 2017 and 2019.

Results: Qualitative data confirmed the quantitative findings. Differences in the experience of EMA between young adults who are currently or were formerly homeless were found to be related to stress or anxiety, interference with daily life, difficulty charging, behavior change, and honesty in responses. Anxiety and depression symptomatology decreased from before to after EMA; however, compliance was not significantly associated with this decrease.

Conclusions: The results point to special considerations when administering EMA to young adults who are currently or were formerly homeless. EMA appears to be slightly more burdensome for young adults who are currently homeless than for those residing in supportive housing, which are nuances to consider in the study design. The lack of a relationship between study compliance and symptomatology suggests low levels of reactivity.

Introduction

Background

Ecological momentary assessment (EMA), also known as the *experience sampling method*, is an intensive, longitudinal, real-time sampling strategy with widespread adoption in public health research and social sciences [1-4]. EMA, which leverages advancements in mobile computing to deliver repeated surveys throughout a specified measurement period, often by using cell phone technology [5], can reduce recall biases and improve the ecological validity and environmental representativeness of the collected data [6]. Given that young adults experiencing homelessness, of which the prevalence is estimated to be 1 in 10 annually in the United States [7], live relatively unstable lives that are highly affected by their immediate environment [8,9] and have high rates of mobile phone technology adoption [10], EMA may be well suited for use with young adults experiencing homelessness [11,12]. In addition, this method may overcome existing limitations in homelessness research that has relied on methods using retrospective reporting, including cross-sectional designs (eg, single point-in-time measurement) [13] and longitudinal study designs with infrequent measurement points (eg, monthly follow-ups) that include issues with attrition [14]. However, although EMA is likely to be a useful method in homelessness research, it is important to better understand the feasibility and acceptability of EMA in this population.

To date, few studies have used EMA to investigate the daily life experiences of young adults who have experienced homelessness, and several studies have shown the feasibility of EMA in this population. Santa Maria et al [15] provided smartphones to 66 young adults who were homeless aged 18 to 25 years to collect EMAs over 21 days and found daily drug use to be predicted by discrimination, pornography use, alcohol use, and urges for substance use and stealing behaviors. In a different study, Tyler et al [16] distributed mobile phones to implement EMA via SMS text messaging with 150 youths who were homeless aged 16 to 22 years over a 30-day period and found that experiencing physical or sexual victimization on a specific day was positively associated with drinking alcohol later that day. Both studies reported high compliance with completing EMA prompts, and the latter study also reported that participants perceived the study to be of low burden [17].

Another important aspect of EMA feasibility is reactivity. Reactivity is understood to be the extent to which the frequency or quality of behavior changes as a result of being monitored [18]. Understanding the potential reactivity is critically important as it suggests that EMA could serve as a possible intervention or manipulation. The literature has generally found low reactivity when using EMA [19-23], including in college and clinical samples [24,25]. Acorda et al [26] found that young people experiencing homelessness were highly receptive to EMA but may have experienced limitations regarding the use of technology and that the repetition of EMA prompts may have affected some behaviors of those participating. However, the extent to which EMA may affect the attitudes or behaviors of young adults experiencing homelessness during a time of identity formation and instability, especially when examining risk behaviors, including sex risk and substance use, has not been well investigated.

Objective

More research is needed to understand the daily experiences of young people who have experienced homelessness, including those who have transitioned into supportive housing, which is a primary intervention being applied to homelessness. Previous work has identified ways in which housed and unhoused young adults differ, including abuse at home [27], which affects mental health [28,29] and substance use [30]. To understand the environmental influences on young adults who have experienced homelessness, it is imperative to examine both those who are currently experiencing homelessness and those who have transitioned from homelessness to supportive housing environments. This study seeks to address this gap by using a mixed methods approach to examine whether there are differences between young adults who are currently (ie, unhoused) versus were formerly homeless (ie, housed in supportive housing) in terms of acceptability, compliance, and reactivity to EMA.

Methods

Study Design

This mixed methods study examines the experiences of EMA in a sample of young adults currently experiencing homelessness and young adults who were formerly homeless who have been placed into supportive housing programs. Specifically, as described in a previously published research protocol paper, young adults participated in a study on health risk behaviors using geographic EMA through a smartphone app that allowed for the collection of time-stamped geographic location data along with EMA behavioral data. Consistent with previous literature [15-17], high compliance with completing EMA prompts (80.2% across the entire sample) over a 1-week study period for the combined sample has already been reported [31]. For this study, we first compared the feedback of housed and unhoused participants regarding their experiences of participating in the EMA week. Responses were then used to examine rates and predictors of EMA compliance and survey responses regarding acceptability and feasibility, comparing those in housing with those who were currently homeless (ie, unhoused). Reactivity was examined using reported anxiety and depression symptomatology before and after the EMA week. Next, we used a qualitative approach to analyze the focus group data of participants who are current or were formerly homeless to better understand their experiences with EMA, which may help explain our quantitative findings.

Participants

Participants (N=231) in transitional living programs or permanent supportive housing (ie, *housed* sample; n=122, 52.8%) and participants not in housing programs (ie, *unhoused* sample; n=109, 47.2%) were enrolled in the study in Greater Los Angeles using stratified convenience sampling. Unhoused participants were recruited via drop-in centers and emergency shelters, including individuals who were explicitly homeless or

unstably housed with temporary living situations that were not reliable beyond 30 days (eg, temporarily crashing with a friend or family member or couch surfing). Participants who consented to the EMA component of the study received up to US $90 in scaled compensation based on response rates and were given the choice of using a study phone with an unlimited data plan or their own smartphone, with an additional US $10 compensation for using their own data plan.

Ethics Approval

All protocols and procedures were approved by the institutional review board at the University of Southern California (review number: UP-16-00046) [31].

Quantitative Component

Overview

Participants were enrolled in a 7-day EMA study comprising questions that asked participants to report on current and previous 2-hour experiences, which were delivered approximately every 2 hours during waking hours using a custom-built app for smartphones using the Android operating system (Google). This study used custom EMA software written by the investigative team. Phones were programmed to only deliver prompts during the waking day, which was determined using the participants' individual estimated sleep and wake times. Prompted surveys asked about physical and social environments, as well as affect and substance use. Participants received an average of 5 EMA prompts per day.

In addition to the EMA prompts, participants completed a daily survey for each EMA day. Daily diaries captured the risk behaviors of the previous day and infrequent behaviors that may be missed by EMAs. Daily survey prompts were scheduled to be delivered at a participant's preferred time but were also available to access via the app at any time during the day to report on the previous day. Daily surveys inquired about participants' social environments, sex behaviors, and substance use.

Before the EMA week, participants completed a baseline interview that took an average of 60 minutes to gather demographic information and data on their histories of homelessness, mental health, and other behaviors. Following the EMA week, participants participated in an exit survey, in which their thoughts and feelings regarding their participation in the EMA study were gathered. The exit surveys lasted approximately 30 minutes. Participants returned the phones and were paid for study participation at the conclusion of the exit survey. The Patient Health Questionnaire-9 was used to assess depression symptomatology [32], and the General Anxiety Disorder-7 was used to assess anxiety symptomatology [33] at both the baseline and exit surveys.

To reduce missed surveys, there were multiple push notifications for both the EMA and daily survey prompts. EMA prompts required a response within 10 minutes after the first prompt, which comprised a chime and vibration. During this 10-minute window, push notifications were sent every 3 minutes. After 10 minutes, the EMA prompt became inaccessible to ensure momentary reporting of the current time and day. Daily surveys

were programmed to send push notifications at 3 time points during the day; however, they could be answered at any point within the day; when answering the daily surveys, participants reported on the prior waking day. The complete study methods are available for further review elsewhere [31] (see Multimedia Appendix 1 for the complete EMA questionnaire and Multimedia Appendix 2 for the daily survey questions).

Analyses

Quantitative analyses in this study included chi-square analysis to compare results by housing status and bivariate linear regressions to predict EMA and daily compliance. Compliance measures the total number of prompts answered out of those received. As the aim of this study was to examine personal factors, such as housing status, rather than artifacts associated with EMA technology, which are associated with compliance, we chose to calculate compliance based on the number of prompts received instead of prompts possible (ie, scheduled). Furthermore, between-subject mixed effects regressions assessed reactivity using the Patient Health Questionnaire-9 [32] and General Anxiety Disorder-7 [33], with random intercepts for each participant.

Qualitative Component

Overview

A total of 4 separate focus groups were conducted to better understand participants' experiences with EMA, each of which occurred within 2 weeks of their last day in the study and lasted approximately 1 hour. Focus groups were chosen as a time-saving way of easily measuring and capturing wide-ranging reactions to EMA. A total of two focus groups included participants recruited from housing programs (one with n=12 transitional living program residents and one with n=6 permanent supportive housing program residents) and 2 focus groups (n=6 and n=7) recruited from youth drop-in centers. Focus group facilitators began by asking participants about their general experiences in the study (eg, "What did you like? What did you not like?"). Additional probing questions included specific aspects of study participation (eg, whether EMA interfered with their daily lives), perceived reactivity to EMA surveys, how accurate they thought their reporting was, level of comfort reporting about sensitive topics such as drugs and alcohol, timing and density of survey prompts, and suggestions for similar studies in the future.

Analyses

Focus group recordings were transcribed and evaluated by 2 independent reviewers, one of whom was the focus group facilitator. Coding and case summaries took a deductive approach using exit survey questions focused on experiences of study participation as a guide (see Multimedia Appendix 3 for the focus group interview guide). Co-coder consensus was achieved through the codevelopment of 4 individual case summaries that summarized each item, including quotations, with 1 for each focus group. Case summaries were then analyzed, first considering housed and unhoused groups separately and then together, to see what might account for and expand upon the differences found.

Results

Quantitative: Exit Interview by Housing Status

Table 1 describes the sample characteristics by housing status, and Table 2 describes the exit surveys and responses, also by housing status. Approximately two-thirds of the participants had an overall positive experience with the study, >90% reported they would be willing to participate in the study again, and approximately 67.5% (156/231) reported that the study took place during a typical week. Approximately 69.7% (161/231) did not feel judged about their sex or drug use, and approximately half of the participants would prefer to use their own phone in a new study provided they owned a phone compatible with EMA technology. As a result of personal choice and incompatibility, only 9% (21/231) used a personal phone

in this study, and significantly more unhoused individuals opted for personal phone use. Housed and unhoused participants also reported statistically significant differences in their experiences of charging the phones, behavior change because of EMA content, being open and honest about EMA survey questions, and whether EMA interfered with their daily life. Compared with those in housing, unhoused participants reported greater difficulty charging their phones ($P=.007$ to overall $P=.02$), greater self-perceived behavior changes in response to EMA ($P=.001$ to overall $P<.001$), that EMA interfered more with their daily life (specific and overall $P<.001$), and more stress or anxiety because of EMA surveys ($P=.008$ to overall $P=.02$), whereas housed participants reported that they were more comfortable answering EMA survey questions openly and honestly (overall $P=.03$).

Table 1. Sample characteristics by housing status (N=231).

Characteristics	Housed (n=122)	Unhoused (n=109)	All participants	*P* value
Age (years), mean (SD)	22.6 (2.4)	21.8 (2.0)	22.2 (2.3)	.007
Gender, n (%)				.06
Male	56 (45.9)	67 (61.5)	123 (53.3)	
Female	46 (37.7)	31 (28.4)	77 (33.3)	
Gender nonconforming, expansive, or transgender	20 (16.4)	11 (10.1)	31 (13.4)	
Sexual minority, n (%)	58 (47.5)	51 (46.8)	109 (47.2)	.91
Race or ethnicity, n (%)				.01
White	11 (9)	10 (9.2)	21 (9.1)	
Black	31 (25.4)	51 (46.8)	82 (35.5)	
Hispanic or Latino	31 (25.4)	17 (15.6)	48 (20.8)	
Biracial or multiracial	36 (29.5)	25 (22.9)	61 (26.4)	
Lifetime homelessness (years), n (%)				.86
<1	45 (36.9)	46 (42.2)	91 (39.4)	
1-2	32 (26.2)	25 (22.9)	57 (24.7)	
3-4	23 (18.9)	19 (17.4)	42 (18.2)	
≥5	22 (18)	19 (17.4)	41 (17.6)	
EMA[a] compliance, mean (SD)	0.80 (0.15)	0.78 (0.19)	0.79 (0.17)	.39
Daily compliance, mean (SD)	0.914 (0.15)	0.906 (0.17)	0.910 (0.16)	.71

[a]EMA: ecological momentary assessment.

Table 2. Exit survey by housing status (N=231).

Survey questions	Housed (n=122)	Unhoused (n=109)	All participants	*P* value
General experience, n (%)				.66
Very negative or somewhat negative	0 (0)	1 (0.9)	1 (0.4)	
Neutral	27 (22.1)	33 (30.3)	60 (25.9)	
Very positive or somewhat positive	55 (45)	65 (59.6)	120 (51.9)	
Difficulty charging, n (%)				.02
Not at all or a little bit	62 (50.8)	61 (55.9)	123 (53.2)	
Somewhat[a]	5 (4)	15 (14.1)	20 (8.7)	
Quite a bit or very much[a]	10 (8.1)	23 (21.1)	33 (14.3)	
Behavioral change, n (%)				<.001
Disagree or strongly disagree[a]	74 (60.7)	36 (33)	110 (47.6)	
Neither agree nor disagree[a]	17 (13.9)	27 (24.8)	44 (19)	
Agree or strongly agree	17 (13.9)	36 (33)	53 (22.9)	
Openness or honesty, n (%)				.03
Disagree or strongly disagree	4 (3.3)	2 (1.8)	6 (2.6)	
Neither agree nor disagree	2 (1.6)	10 (9.1)	12 (5.2)	
Agree or strongly agree	102 (83.6)	87 (79.8)	189 (81.8)	
Interference with life, n (%)				<.001
Not at all or a little bit	75 (61.5)	45 (41.3)	120 (51.9)	
Somewhat[a]	23 (18.9)	22 (20.2)	45 (19.5)	
Quite a bit or very much[a]	9 (7.4)	32 (29.4)	41 (17.7)	
Stress or anxiety, n (%)				.02
Not at all or a little bit	96 (78.7)	74 (67.9)	170 (73.6)	
Somewhat[a]	8 (6.6)	12 (11)	20 (8.7)	
Quite a bit or very much[a]	4 (3.3)	13 (11.9)	17 (7.4)	
Willingness to follow up, n (%)				.07
No	2 (1.6)	7 (6.4)	9 (3.9)	
Yes	106 (86.9)	92 (84.4)	198 (85.7)	
Feeling judged, n (%)				.99
Disagree or strongly disagree	84 (68.9)	77 (70.6)	161 (69.7)	
Neither agree nor disagree	10 (8.1)	9 (8.2)	19 (8.2)	
Agree or strongly agree	14 (11.5)	13 (11.9)	27 (11.7)	
Willingness to participate again, n (%)				.48
Neither agree nor disagree	7 (5.7)	9 (8.2)	16 (6.9)	
Agree or strongly agree	100 (81.9)	89 (81.7)	189 (81.8)	
Phone preference (new study), n (%)				.80
Study phone	57 (46.7)	54 (49.5)	111 (48.1)	
Personal phone	51 (41.8)	45 (41.3)	96 (41.6)	
Phone used this study, n (%)				.07
Personal phone	7 (5.7)	14 (12.8)	21 (90.9)	
Study phone	101 (82.8)	85 (77.9)	186 (80.2)	

XSL•FO
RenderX

Survey questions	Housed (n=122)	Unhoused (n=109)	All participants	*P* value
Typical week, n (%)				.14
No, this week was not typical	22 (18)	29 (26.6)	51 (41.8)	
Yes, this was a typical week	86 (70.5)	70 (64.2)	156 (67.5)	

[a]Combined for specific *P* values.

Quantitative: Delivery and Compliance

Out of a theoretical maximum of 20,076 prompts, 17,944 (89.38%) prompts were scheduled for delivery by the custom software. The discrepancy of the 10.62% (2132/20,076) of prompts may be explained by hardware issues (eg, low battery), software issues (eg, app crashing), or schedule timing (ie, where a participant was enrolled at midday and would not have received earlier prompts). Of the scheduled prompts, approximately 39.57% (7101/17,944) were not delivered as the app detected that the prompt time was within sleep parameters specified by the participant. Of the 17,944 prompts, 679 (3.78%) scheduled prompts were not delivered because of Android system features designed to conserve battery life and memory, and 138 (0.77%) prompts were not delivered as the phone was intentionally turned off. The remaining undelivered scheduled prompts, which was an average of 46, were missing because of unknown software or hardware errors. In all, of the 17,944 prompts, the participants received 9980 (55.62%) prompts and completed 8001 surveys upon answering the prompts (8001/9980, 80.17% EMA completion). Participants failed to answer 18.24% (1820/9980) of prompts and answered but did not complete 1.59% (159/9980) of surveys. Participants took, on average, 97 (SD 70, range 19-599) seconds to complete the EMA surveys and completed 13.8 (SD 6.4, range 0-25) questions per EMA survey (see Multimedia Appendix 1 for the EMA survey questions and Figure 1 for an example screenshot of the app).

Figure 1. Screenshot of the ecological momentary assessment presented to participants.
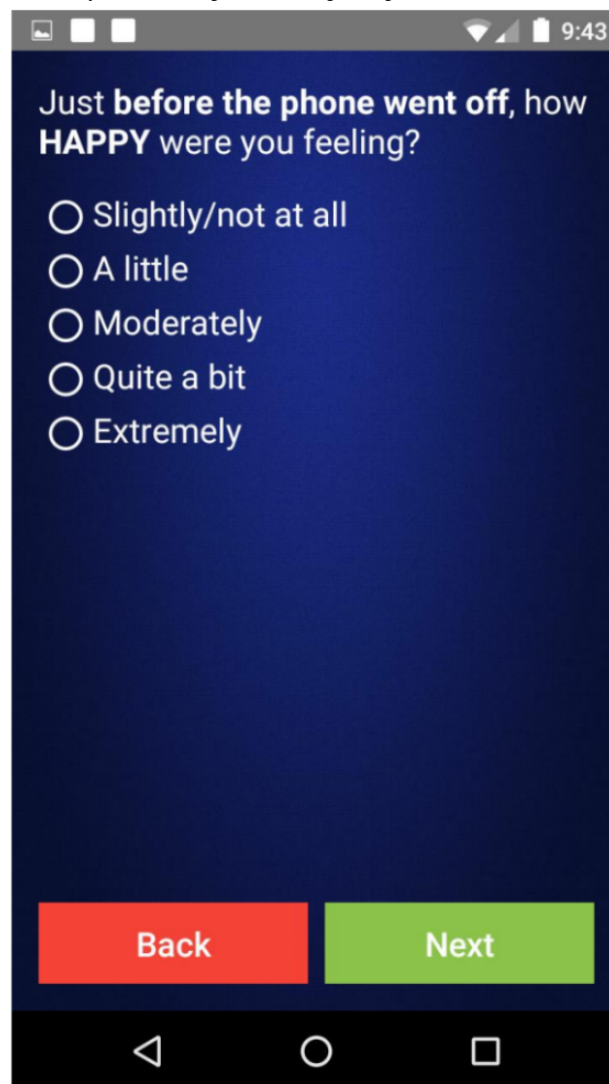
Approximately 3% (7/231) of participants did not receive 2.93% (49/1673) of daily survey prompts, and an additional 6.69% (112/1673) of daily survey prompts were not delivered, out of the theoretical maximum number of prompts, because of possible hardware, software, or schedule timing issues. Of the 1512 scheduled daily survey prompts, 5 (0.33%) were not delivered because of Android system battery conservation issues, and 5 (0.33%) were not delivered as the phone was turned off by the participant. Out of 1502 prompts that were received, participants answered and completed 1376 (91.61% daily compliance) daily prompts; of the 1512 surveys, participants failed to fully complete 6 (0.4%) surveys. Participants took, on average, 89 (SD 61, range 14-570) seconds to complete daily surveys, and they completed 12.2 (SD 5.1, range 0-45) questions per daily survey (see Multimedia Appendix 2 for the daily survey questions).

The results of the analyses of compliance are presented in Table 3. Neither daily compliance nor EMA compliance was associated with housing status ($t_1=-0.38$, $P=.71$ and $t_1=-0.86$, $P=.39$, respectively). Only 9.1% (21/231) of participants chose to complete the study on their own phone, with no difference in compliance between those who used a personal versus a study phone ($t_1=1.15$, $P=.25$ and $t_1=0.74$, $P=.46$, respectively). Compared with those reporting negative or neutral experiences, participants who reported a *very positive* or *somewhat positive* experience with the study had 6% (SE 0.03%) greater EMA compliance ($t_1=2.49$; $P=.01$). Furthermore, those reporting *not at all* or *a little bit* of difficulty charging their device had 6% (SE 0.03%) greater EMA compliance than those reporting more difficulty charging their device ($t_1=2.41$, $P=.02$).

Daily survey compliance was not associated with general experience with the study ($t_1=0.97$; $P=.33$), nor was participants' self-report of honesty regarding survey responses associated with EMA or daily compliance ($t_1=0.92$, $P=.36$ and $t_1=1.23$, $P=.22$, respectively). Compared with those reporting somewhat or greater interference with the study protocol in their lives, participants who reported not at all to a little bit of interference had 8% (SE 0.02%) greater EMA compliance ($t_1=-3.83$; $P<.001$) and 6% (SE 0.02%) greater daily compliance ($t_1=-3.42$; $P=.001$). Similarly, participants experiencing little or no stress or anxiety from surveys had 7% (SE 0.03%) greater EMA compliance ($t_1=2.62$; $P=.009$) and 5% (SE 0.03%) daily compliance ($t_1=2.05$; $P=.04$) than those experiencing greater stress or anxiety from surveys. Compared with those who indicated feeling judged by surveys, participants who did not endorse feeling judged by the surveys had 6% (SE 0.03%) greater EMA compliance ($t_1=2.38$; $P=.02$) and 9% (SE 0.02%) greater daily compliance ($t_1=4.02$; $P<.001$). Participants who reported having a typical week had 8% (SE 0.02%) greater EMA compliance than those who reported having an atypical week ($t_1=3.58$; $P<.001$); however, those with typical weeks only had marginally greater daily compliance ($t_1=1.79$; $P=.08$). Willingness to participate again was not associated with EMA or daily compliance ($t_1=0.34$, $P=.74$ and $t_1=-0.05$, $P=.96$, respectively).

Table 3. Bivariate linear regressions of compliance (N=231).

Characteristics	Sample size, n (%)	EMA[a] compliance			Daily compliance		
		Margin (SE)	β[b] (SE; 95% CI)	P value	Margin (SE)	β[b] (SE; 95% CI)	P value
Housing status							
Unhoused (reference)	231 (100)	0.78 (0.02)	N/A[c]	N/A	0.91 (0.02)	N/A	N/A
Housed	231 (100)	0.80 (0.02)	.02 (.02; −0.03 to 0.64)	.39	0.91 (0.01)	.008 (.02; −0.03 to 0.05)	.71
Phone type							
Personal phone (reference)	207 (89.6)	0.78 (0.03)	N/A	N/A	0.89 (0.03)	N/A	N/A
Study phone	207 (89.6)	0.81 (0.01)	.03 (.04; −0.05 to 0.10)	.46	0.93 (0.01)	.04 (.03; −0.03 to 0.10)	.25
Study experience							
Negative or neutral (reference)	181 (78.4)	0.76 (0.02)	N/A	N/A	0.93 (0.01)	N/A	N/A
Positive	181 (78.4)	0.82 (0.01)	.06 (.03; 0.01 to 0.11)	.01	0.91 (0.02)	.02 (.02; −0.02 to 0.06)	.33
Difficulty charging							
None or a little (reference)	176 (76.2)	0.83 (0.01)	N/A	N/A	0.94 (0.01)	N/A	N/A
Somewhat or yes	176 (76.2)	0.76 (0.02)	−.06 (.03; −0.11 to −0.01)	.02	0.90 (0.02)	−.04 (.02; −0.08 to 0.004)	.07
Behavior change							
Yes or neutral (reference)	207 (89.6)	0.79 (0.02)	N/A	N/A	0.92 (0.01)	N/A	N/A
No	207 (89.6)	0.82 (0.02)	.04 (.02; −0.004 to 0.08)	.08	0.92 (0.01)	−.0008 (.02; −0.04 to 0.04)	.97
Openness or honesty							
No or neutral (reference)	207 (89.6)	0.77 (0.04)	N/A	N/A	0.88 (0.03)	N/A	N/A
Yes	207 (89.6)	0.80 (0.01)	.04 (.04; −0.04 to 0.11)	.36	0.93 (0.01)	.04 (.03; −0.03 to 0.11)	.22
Interference with life							
None or a little (reference)	206 (89.2)	0.84 (0.01)	N/A	N/A	0.95 (0.01)	N/A	N/A
Somewhat or yes	206 (89.2)	0.76 (0.02)	−.08 (.02; −0.13 to −0.04)	<.001	0.89 (0.01)	−.07 (.02; −0.10 to −0.03)	.001
Stress or anxiety							
None or a little (reference)	207 (89.6)	0.82 (0.01)	N/A	N/A	0.93 (0.01)	N/A	N/A
Somewhat or yes	207 (89.6)	0.74 (0.03)	−.07 (.03; −0.13 to −0.02)	.009	0.88 (0.02)	−.05 (.03; −0.10 to −0.002)	.04
Feeling judged							
Yes or neutral (reference)	207 (89.6)	0.76 (0.02)	N/A	N/A	0.85 (0.02)	N/A	N/A
No	207 (89.6)	0.82 (0.01)	.06 (.03; 0.01 to 0.11)	.02	0.94 (0.01)	.09 (.02; 0.05 to 0.14)	<.001
Typical week							
No (reference)	207 (89.6)	0.74 (0.02)	N/A	N/A	0.89 (0.02)	N/A	N/A
Yes	207 (89.6)	0.83 (0.01)	.09 (.02; 0.04 to 0.14)	<.001	0.93 (0.01)	.04 (.02; −0.004 to 0.08)	.08
Willingness to participate again							
Neither agree nor disagree (reference)	205 (88.7)	0.79 (0.04)	N/A	N/A	0.93 (0.04)	N/A	N/A

XSL•FO
RenderX

Characteristics	Sample size, n (%)	EMA[a] compliance			Daily compliance		
		Margin (SE)	β[b] (SE; 95% CI)	*P* value	Margin (SE)	β[b] (SE; 95% CI)	*P* value
Yes, agree	205 (88.7)	0.81 (0.01)	.01 (.04; −0.07 to 0.10)	.74	0.92 (0.01)	−.002 (.04; −0.07 to 0.07)	.96

[a]EMA: ecological momentary assessment.

[b]Indicates percentage change in compliance.

[c]N/A: not applicable.

Quantitative: Reactivity Analyses

Table 4 displays the results from the mixed effects models to examine reactivity to EMA participation, specifically regarding anxiety and depression symptomatology. Both anxiety and depression scores decreased from baseline to follow-up (β=−1.77, *P*<.001 and β=−1.10, *P*=.03, respectively). However, no significant effects for EMA compliance were detected for either anxiety or depression; thus, the decrease in symptomatology was not associated with compliance. Both models controlled for age, gender, sexual orientation, race and ethnicity, and housing status.

Table 4. Mixed effects regressions for reactivity (N=231).

Characteristics	Anxiety (GAD-7[a])		Depression (PHQ-9[b])	
	β (95% CI)	*P* value	β (95% CI)	*P* value
Time point (*j*)	−1.79 (−2.62 to −0.96)	<.001	−1.08 (−2.07 to −0.09)	.03
Compliance (EMA[c])	1.11 (−2.88 to 5.10)	.59	.62 (−3.62 to 4.86)	.78
Housing status (housed)	−.02 (−1.49 to 1.45)	.98	−.58 (−2.14 to 0.98)	.47
Age (years)	.14 (−0.17 to 0.46)	.37	.01 (−0.32 to 0.35)	.94
Gender (reference: male)				
Female	.21 (−1.34 to 1.77)	.79	−.12 (−1.77 to 1.54)	.89
Gender nonconforming, expansive, or transgender	2.24 (−0.11 to 4.59)	.06	3.77 (1.28 to 6.25)	.003
Sexual minority (reference: heterosexual)	3.63 (2.17 to 5.09)	<.001	3.64 (2.10 to 5.19)	<.001
Race (Black)	−.99 (−2.59 to 0.60)	.22	−1.27 (−2.96 to 0.42)	.14
Hispanic	1.11 (−0.48 to 2.71)	.17	1.02 (−0.67 to 2.71)	.24

[a]GAD-7: General Anxiety Disorder-7.

[b]PHQ-9: Patient Health Questionnaire-9.

[c]EMA: ecological momentary assessment.

Qualitative Findings

Overview

The qualitative findings that were generated independently of the quantitative findings were categorized under the 5 main emergent themes. The first theme explains how participants felt the study design *increased mindfulness and reflection*, whereas the second theme captures those negative instances when the participation in the study resulted in *causing stress and anxiety*. The third theme discusses the ways in which study participation incited behavior change, whereas the fourth theme addresses participants *responding honestly* to questions. The final theme captures participant suggestions about future study designs. We note that a comparative analysis of the housed and unhoused samples indicated that these themes apply to both groups, as shown in Table 5.

Table 5. Qualitative findings by housing status.

Themes	Housed focus groups (n=18)	Unhoused focus groups (n=13)
Increased mindfulness and reflection	• "You get to know yourself. Like how many times do I do these things a day? Who am I around every day? What am I doing every day? So, it gave me insight on who you are and what you do every day. Because sometimes we'll just do things. And we don't keep track of those things. They kind of keep you on track a little bit." [SP[a]102] • "Some stuff I forgot about, that...It just opened my mind more. I was, 'Okay, I need to start paying more attention to that. I need to start paying more attention to this. I am around certain people who do stuff like this.'" [SP209]	• "It was the shit to me. It actually calmed me down on most occasions." [SP305] • "It makes you feel good. You want to get up and answer the questions that, how you feel, how you feel going out today, and shit like that. It made me feel good, honestly." [SP306] • "Yeah 'cause I forgot to do my chores and as soon as that survey come on, and be like, 'Oh snap. I forgot to do my chore.'...But then I'm right back on it." [SP403] • "It also gave me a good understanding of how when I hang out with these certain people, yeah, I am smoking more. And if I hang out with this certain people, I am drinking more...Gonna be like, okay dude, I'm hanging out with you but just because we're hanging out doesn't mean we have to drink. You know? If they're drinking it's their choice. Alright man, I'm noticing I can't be hanging out with you every day. You're drinking every day, I'm hanging out with you every day, I'm probably gonna be drinking every day." [SP406]
Causing stress and anxiety	• "Well, it was irritating, like once or twice. I thought it was going to ask different questions. I didn't know about the repetition thing." [SP106] • "Sometimes it would annoy me, like, 'How do you feel?' Like, 'Oh, I feel annoyed now.'" [SP207]	• "I got paranoid sometimes, like if I was hanging out with these people and it was asking me, 'who are you with?' 'Did you do any drugs with them?'...It was just about the whole street thing; it feels like snitching." [SP301] • "I got annoyed sometimes. If I was going through something and the survey went off, I just didn't want to answer it sometimes." [SP302] • "I did [feel uncomfortable] at first. I thought it was like the Feds or something. I was like oh shit, I'm not gonna lie, I was doing all types of lies though. A few of the surveys I feel like I failed them or something. I don't know. That's just how my mentality think. I felt like the phone was recording. Something. On everything. Yeah, I felt like the Feds was watching. And taking video at the same time. I let the phone die for like a day and a half...I'm like I don't know, I'm gonna keep the phone off. I was just annoyed about it because like they know oh really alerted when I got a text that 'we see your phone hasn't been charged.' [laughs] You guys are watching me! I know what you can do with technology. It's a simple program, there's no telling what's written in that program that I don't see." [SP403]
Inciting behavior change	• "I'm not going to drink today, because they're going to ask me how many drinks I drank." [SP102] • "I had kind of like a Pavlov's dog affect, where every time it went off, I'm like, 'Ooh, I could really use some alcohol right now.' I'm like, 'Ooh, this is reminding me that alcohol is not a great option, but it is an option.' [SP206]	• "Have you smoked yet today? Okay? I haven't, mostly because I haven't got my weed so hold on, thanks for letting me know I have to go get weed. That's what I'm saying. Reminding me to go get my marijuana when I run out." [SP305] • "On the alcohol and drink question I went like...the repetitive asking you another question and just mentioning a drink, made me want to drink. I never had that many drinks in a week." [SP403]

Themes	Housed focus groups (n=18)	Unhoused focus groups (n=13)
Responding honestly	• "I tried to be as honest as possible. Because I knew that it was a study. So I tried to be as honest...Because you guys are going to look at it and try to get real answers from people. It was hard. But I tried to be honest about it." [SP102] • "The thing about surveys too, is it's a non-judgemental environment that you're telling your information to." [SP105] • "Yeah, I feel like if anything it makes me feel more comfortable, because the phone, you don't have to give a fuck...I'm going to be real today, and I'm going to be real to myself. It's kind of like a diary if you think about it. That survey was like a diary for a week...If you have time for that." [SP205] • "To be honest, I lied on a few questions...I lied because- It just reminded me of what a whore I am, because like...so it would be like 'How many times did you have sex today?' And stuff like that, I'm like 'Ooh.' [SP202]	• "...it was just about the whole street thing, it feels like snitching. Yep." [SP301] • "I say it depends on the mood of how people will feel when I'm doing the surveys." [SP306] • "A lot of them I tried to be honest some, some of them I kind of like, I don't think I put the right answers, like it would say, where you at, and sometimes I'd put like, or what are you doing, and I'd say hanging out, but really I was eating, still hanging out a little bit, so kind of lie about that, I need to be more honest, cause you know, I don't want them to know exactly what I'm doing." [SP404] • "I thought it was like the Feds or something. I was like oh shit, I'm not gonna lie, I was doing all types of lies though." [SP403]
Suggestions for future studies	• "I think [using my own phone] was more convenient. It was a lot easier than having to have a second phone. Losing track of it." [SP105] • "I thought it was interesting. It was cool. I think I prefer it to be on my actual phone than another phone. Because it was kind of hard to keep up with it." [SP104] • "It felt extra for me but that's because...I'm really bad about keeping my regular phone on me, so I said I was a bad millennial. I'm a bad millennial." [SP204] • "It's hard to remember to carry two phones." [SP208] • "But even if I went from my bedroom to the dining room, and thirty minutes went by and I was like, 'I have to go get my phone.' Then I saw that I missed one survey." [SP204] • "I feel like I have bigger problems than carrying two phones. It wasn't the hardest thing in my life, but sometimes I would forget it and be like, 'Aw, crap.' That's it." [SP207] • "It was hard for me. I was literally driving. I always keep my phone away. I'm over there trying to jump in the purse like: Where is this phone? And then, another thing, my neighbor...My neighbor [who was also in the study] always comes over...And my other neighbor...So he'd be like: 'Who's phone is who's? Who got that...' And I'd be like: 'Do you have my phone? That's my phone?' How did we know it was our phone?...It took me about a day, or maybe two, to actually [get used to it]." [SP104]	• "[Borrowing] the phone...actually helped me with my daily life, bro. Because I didn't have a phone at the time, so it helped me to get phone calls." [SP404] • "Oh yeah, me too. Google maps." [SP408] • "I hadn't had a phone since that one. I do want a phone for another week though." [SP404] • "Because it would ask have you did this in the past 2 hours and I might forget. Or how many cigarettes, how many times do you think you used tobacco products. Like, I don't know, let me count the cigarettes in my pack right now...I might forget and I just smoked a cigarette 5 minutes before it happened. So more every time you smoke a cigarette, log it in the phone. Tally it." [SP402]

^aSP: study participant

Increased Self-reflection and Self-awareness

Self-reflection and awareness while participating in EMA were common points of discussion in the focus groups. In fact, this idea came up in all 4 focus groups. Many participants noted increased self-awareness regarding how their own thoughts, feelings, and social or physical contexts influenced their engagement in protective or risky behaviors.

For example, several participants noted that the EMA helped them stay on track over the course of the day, including triggering better awareness of time, what has been accomplished, and what has yet to be done. A participant discussed being able to keep track of what they were up to:

You get to know yourself. Like how many times do I do these things a day? Who am I around every day? What am I doing every day? So, it gave me insight on who you are and what you do every day. Because sometimes we'll just do things. And we don't keep

track of those things. They kind of keep you on track a little bit. [Study participant (SP) 102, housed]

Similarly, another participant specifically talked about how EMA kept them productive and on top of their chores, saying the following:

Yeah 'cause I forgot to do my chores and as soon as that survey come on, and be like, "Oh snap. I forgot to do my chore."...But then I'm right back on it. [SP403, unhoused]

Several participants realized that they were using drugs and alcohol more than they thought by logging their substance use every 2 hours:

Oh my gosh, I was shocked how many times I actually put down how much I drank alcohol, and truthfully, I have been trying to stop. [SP408, unhoused]

Furthermore, substance use was also linked to social and physical environments for some:

It also gave me a good understanding of how when I hang out with these certain people, yeah, I am smoking more. And if I hang out with certain people, I am drinking more...Gonna be like, okay dude, I'm hanging out with you but just because we're hanging out doesn't mean we have to drink. You know? If they're drinking it's their choice. Alright man, I'm noticing I can't be hanging out with you every day. You're drinking every day, I'm hanging out with you every day, I'm probably gonna be drinking every day. [SP406, unhoused]

Questions about drug use and sexual activity were often considered sensitive topics—topics that others might not ask the participants about because of their sensitive nature. Some participants felt that the EMA prompts provided a safe and comfortable opportunity to check in:

I think you get to know more about yourself by...You're being asked questions that a whole lot of people wouldn't ask you. So I found it interesting for a device or a system, that's not human, obviously, to ask you those questions. I was like: Wow. They were pretty insightful. Even if they were like the same old questions over and over and over. "How much drugs do you take?" Or "Who do you sleep with?". It's information that you want to keep private. You don't want to share it with anybody. And also, it's kind of a comfortable feeling too. [SP105, housed]

In addition, the *check in* and ability to confide in the surveys seemed to provide a sense of calm for participants: such as who stated:

It took me a little break, too. That was my little break, with the alarm going off. I was like oh shit, I'm on my little couple seconds break. [SP403, unhoused]

Self-reflection Negative Case Analysis: Causing Stress and Anxiety

Although self-reflection and self-awareness were common in focus group discussions, which seemed to promote a sense of calmness among participants, aspects of the EMA protocol also seemed to trigger stress, anxiety, and paranoia, particularly in response to location tracking. Approximately 16% (37/231) of participants reported EMA caused *somewhat* or *quite a bit* of stress or anxiety in the exit survey (Table 2). Although some participants liked the personal nature of the questions, allowing them to check in with themselves regarding personal behaviors that were otherwise likely to go unchecked, others felt that the questions were too personal. In fact, a participant discussed paranoia as a result of the prompts:

I got paranoid sometimes, like if I was hanging out with these people and it was asking me, "who are you with?" "Did you do any drugs with?"...It was just about the whole street thing; it feels like snitching. [SP301, unhoused]

Similarly, another participant spoke of the discomfort they felt regarding the nature of the questions and location monitoring:

I did [feel uncomfortable] at first. I thought it was like the Feds or something. I was like oh shit, I'm not gonna lie, I was doing all types of lies...I felt like the Feds was watching. And taking video at the same time. I let the phone die for like a day and a half...I'm like I don't know, I'm gonna keep the phone off. I was just annoyed about it them because like they know oh really alerted when I got a text that "we see your phone hasn't been charged." [laughs] You guys are watching me! I know what you can do with technology. It's a simple program, there's no telling what's written in that program that I don't see. [SP403, unhoused]

Fears regarding the personal nature of questions and location tracking, such as those discussed by participants SP301 and SP403, who were unhoused and both enrolled via drop-in centers, were more common among young adult participants who were actively homeless (ie, focus groups 3 and 4 [see Table 5 for more quotes by focus group and housing status]).

Further stress appeared to arise from the repetitive nature of the prompting schedule, with EMA prompts approximately 2 hours apart with the same array of questions. A participant commented on this by saying the following:

Well, it was irritating, like once or twice. I thought it was going to ask different questions. I didn't know about the repetition thing. [SP106, housed]

Some participants felt that the prompting occurred too frequently and reported that the prompt changed their mood and what they reported as the prompt itself annoyed them:

Sometimes it would annoy me, like, "How do you feel?" Like, oh, I feel annoyed now [SP207, housed]

A similar response was recorded by another participant, who commented the following:

Damn, this alarm's going off and now I'm irritated. Like I was doing fine before but now I'm irritated as fuck. [SP202, housed]

An additional point about the prompting scheme focused on the concept of negativity in survey items, as a few participants noted that the surveys did not ask about positive things in their lives, such as work, school, or other productive aspects of their lives. A participant commented the following:

It seemed to always be focused on like, what did you do that wasn't productive today. How much weed did you smoke how many did you smoke, how much did you drink, did you fuck anybody for drugs, what time did you wake up, did you even go to sleep tonight—it didn't ask anything like I don't know did you have a good day or did you...Did anything positive happen to you today. It was just all focused on, I'm not going to say negativity...I guess how often homeless youth use drugs and trade that for sex or whatever else. Being on the streets, just assuming being on the streets that's all that you do. [SP403, unhoused]

Despite these concerns, the participants seemed to become used to the prompting schedule over the course of the week.

Specifically, a participant discussed their experience of becoming familiar with the surveys, which made the experience easier:

> I got used to it...At first it was kind of slow, for me. So it'd take a while, maybe five or ten minutes. I'm taking forever on this thing until eventually it was two minutes. And after a while it wasn't really bothering me. [SP404, unhoused]

Considering the demands of the prompting schedule, participants had to juggle their study responsibilities with the responsibilities of their daily lives. Remaining stresses that resulted from study participation focused on having to comply (ie, answer a certain number of prompts) to get paid; needing to keep track of their study phones and answer surveys when at work or school; and not being comfortable with answering in front of others, such as SP408 when in the company of their case manager:

> I know I didn't want to miss one because you're not gonna get your money. I want to get to a good percentage, high percentage [SP404, unhoused]

> If I was working or if I was at school...I want to get the survey done, but I really have to pay attention to what I'm doing. So it's like let me just get it over with [SP103, housed]

> A few times it ringed off when I was in a meeting with my case manager. I was like "oh...no." I couldn't just pull out my phone and start doing a survey in front of my case manager. I was like, why you gotta hit at this hour. [SP408, unhoused]

On occasion, these demands resulted in participants clicking through the surveys. A participant described this as "going into default":

> I did find myself going into a default and then just changing it if it was different because my mood was fairly stable throughout the day and I'd be like, "This one, this one, this one, this one." Then I'm like, "Eh, this one's actually over here now." I definitely would go into a default but then I'd adjust it. [SP202, housed]

Inciting Behavior Change

Approximately 22.9% (53/231) of the SPs thought that their behaviors had changed because of study participation. Most notably, behavior change was focused on drug and/or alcohol use and more often reported by unhoused participants than those residing in supportive housing. On some occasions, EMA promoted positive change, such as the intention to consume less of a substance, such as with one participant's alcohol consumption:

> I'm not going to drink today, because they're going to ask me how many drinks I drank. [SP102, housed]

However, on the other hand, prompting regarding drug or alcohol consumption also had negative impacts, resulting in cravings or the desire to consume. A participant explains the effect as follows:

> I felt a little like after a day or two, I had kind of like a Pavlov's dog effect, where every time it went off, I'm like, "Ooh, I could really use some alcohol right now." I'm like, "Ooh, this is reminding me that alcohol is not a great option, but it is an option." [SP206, housed]

Similarly, another participant stated the following:

> On the alcohol and drink question...the repetitive asking you another question and just mentioning a drink made me want to drink...It kept asking me how many drinks I have had and I'm like, none, but I don't know, I kind of want one now, shit. [SP403, unhoused]

In addition, the subject matter, combined with the annoyance of the prompting schedule, brought on an increased desire to use:

> Especially since it was that ringing and a little bit of an irritant, and it's like, "Ooh, a drink would be really nice to kind of just not deal with this right now" [SP206, housed]

Responding Honestly

Across the 4 focus groups, the participants discussed honesty in their survey responses. Although 81.8% (189/231) of focus group participants agreed with being open and honest, others expressed greater distrust and, therefore, less honesty. Some participants oscillated between providing an honest picture of their day and, at times, lying. One of the participants provides a good example of the latter:

> I should say...I was just trying to be as honest as possible. But some days you just don't feel like explaining yourself or completely answering. [SP103, housed]

Here, fluctuations in fatigue and social engagement influenced honesty in survey responses.

The occasions where participants reported dishonesty, which occurred more often among unhoused participants, were largely related to the personal nature of the questions. One of the participants noted that the more specific and personal the questions got, the harder they were to answer honestly. This participant continued by providing a specific example:

> It was asking me to put a nickname down for someone you had sex with, I was like, I don't really want to do that. [SP103, housed]

Furthermore, another participant described their experience with honest survey responses as follows:

> To be honest, I lied on a few questions...I lied because it just reminded me of what a whore I am. It would be like "How many times did you have sex today?" And stuff like that, I'm like it's clocking me. I had the times I was cheating on my boyfriend with this boy, and you know how it asks you like the five people that you hang out with mostly or whatever? Then it would be like, "Who were you with?" I'd be like, "Ooh." [SP202, housed]

Despite these circumstances, many participants discussed why they chose to respond honestly to the survey items. A participant shared their respect for the study by stating the following:

> *I tried to be as honest as possible. Because I knew that it was a study. So I tried to be as honest...Because you guys are going to look at it and try to get real answers from people. It was hard. But I tried to be honest about it.* [SP102, housed]

Others noted that responding via the phone provided some comfort:

> *The thing about surveys too, is it's a non-judgmental environment that you're telling your information to.* [SP105, housed]

Another participant also had similar feelings about using their phone to answer the surveys honestly:

> *Yeah, I feel like if anything it makes me feel more comfortable, because the phone, you don't have to give a fuck...I'm going to be real today, and I'm going to be real to myself. It's kind of like a diary if you think about it. That survey was like a diary for a week.* [SP205, housed]

Suggestions for Future Studies: Technology

The largest perk for many participants in the study without a phone of their own was receiving a phone with a full data plan for the whole week, which most participants chose to do. Only 9.1% (21/231) of participants used their own phones for the study. Participants from focus group 4 (unhoused) discussed the following:

> *[Borrowing] the phone...actually helped me with my daily life, bro. Because I didn't have a phone at the time, so it helped me to get phone calls.* [SP404]
>
> *Oh yeah, me too. Google maps.* [SP408]
>
> *I hadn't had a phone since that one. I do want a phone for another week though.* [SP404]

Some others with their own personal devices preferred using their own phones rather than needing to keep track of 2 phones for the duration of the study:

> *I think [using my own phone] was more convenient. It was a lot easier than having to have a second phone. Losing track of it.* [SP105, housed]
>
> *I thought it was interesting. It was cool. I think I prefer it to be on my actual phone than another phone. Because it was kind of hard to keep up with it.* [SP104, housed]

In addition, some with their own personal phones chose to borrow a study phone to keep the study separate from their personal lives or because the software was incompatible with their phones. This, most often, resulted in difficulty keeping track of the 2 phones, as discussed in focus group 2 (housed):

> *It felt extra for me but that's because...I'm really bad about keeping my regular phone on me, so I said I was a bad millennial. I'm a bad millennial.* [SP204]
>
> *It's hard to remember to carry two phones.* [SP208]

> *But even if I went from my bedroom to the dining room, and thirty minutes went by and I was like, "I have to go get my phone." Then I saw that I missed one survey.* [SP204]
>
> *I feel like I have bigger problems than carrying two phones. It wasn't the hardest thing in my life, but sometimes I would forget it and be like, "Aw, crap." That's it.* [SP207]

Another discussion point stemming from this issue occurred in friend groups in which multiple people were enrolled in the study at the same time. One of the participants mentioned the difficulty of keeping track of their phone when interacting with others also in the study:

> *It was hard for me. I was literally driving. I always keep my phone away. I'm over there trying to jump in the purse like: Where is this phone? And then, another thing, my neighbor...My neighbor [who was also in the study] always comes over...And my other neighbor...So he'd be like: "Who's phone is who's? Who got that..." And I'd be like: "Do you have my phone? That's my phone?" How did we know it was our phone?...It took me about a day, or maybe two, to actually [get used to it].* [SP104, housed]

To this point, another participant in an unhoused focus group offered a potential way of distinguishing study phones in the same friend group, recommending the following:

> *Maybe have a choice of different ringtones, because when you're in a group of people and the ringtone goes off, but everyone thinks it's theirs...Everyone's like, "Hold on, I think it's mine."*

Finally, and perhaps most notably, SPs recognized the value of EMA as an intervention. Participants wanted more, such as prompts to support them in cutting back on their substance use, with many participants noting the benefits of study participation in terms of mindfulness:

> *I feel like sometimes you're doing stuff and then...Like, say somebody says yes, to something on the list. Like some serious drugs...Meth or something like that. Which you know is a serious drug. You can tell them like: "Maybe you shouldn't do it." Or something like that. I'm not sure about the whole thing. I'd have to sit down and go over it myself. If I was designing the app or something.* [SP101, housed]

Mixed Methods

In comparing the quantitative and qualitative arms of this study via a convergent parallel design, we found qualitative responses from both housed and unhoused participants to confirm the quantitative findings (see Table 6 for an integration of the quantitative and qualitative findings). The findings regarding the previously discussed difficulties with charging devices were found to be convergent. Qualitative findings regarding study-induced stress and anxiety, as well as interference with daily life, provided additional contextual information to the quantitative findings, offering an expansion in the interpretation of results. As previously stated, unhoused participants self-reported greater, statistically significant study-induced

stress and anxiety. However, the qualitative findings highlight that housed individuals, at times, also noted stress and anxiety related to study participation; however, this stress and anxiety seemed to be contextually different from that experienced by unhoused participants, which was often connected to paranoia, fears of snitching, and being *watched by the Feds*. The latter stress and anxiety could be directly related to street culture and economy. Similarly, quantitative findings showed increased

reporting of study interference in daily life among unhoused individuals; however, qualitative findings showed that both unhoused and housed participants noted interference. Housed participants most often talked about interference in terms of school or work responsibilities, whereas those unhoused discussed needing to answer to make the money for survey compliance, which made it more likely to interfere with what they were doing.

Table 6. Results of convergent parallel design based on housing status.

Quantitative findings	Qualitative findings	Merged findings outcome
Compared with housed participants, unhoused participants were more likely to report that the *study caused them stress or anxiety* (*P*=.02).	Although both housed and unhoused participants reported stress accompanying the study, unhoused participants discussed stress related to paranoia and snitching.	Confirmatory; expansion
The study was reported to have *interfered with daily life* among those unhoused compared with those in housing (*P*<.001).	Although surveys appeared at inopportune times for both housed (eg, while at work or school) and unhoused (eg, while visiting with a case manager) participants, unhoused participants discussed more stress regarding needing to answer to make the money for survey compliance, which made it more likely to interfere with what they were doing.	Confirmatory; expansion
Unhouse participants reported greater *difficulty charging* their devices (*P*=.02).	Housed participants were more likely to be in locations with outlets. Charging had to be sought out by unhoused participants.	Confirmatory; convergent
More unhoused participants reported that the study caused *changes in their behavior* (*P*<.001) than those in housing.	Although both housed and unhoused participants noted increased awareness of substance use related to substance use questions, unhoused participants more often reported increased substance use, whereas housed participants seemed to mention the awareness of substance use and trended toward a reduction in their use.	Confirmatory; complementarity
Compared with those in housing, unhoused participants reported being less likely to *be open or honest* when answering survey items (*P*=.03).	Housed participants seemed to feel more comfortable with honest responses, whereas unhoused participants again noted paranoia and fear of snitching.	Confirmatory; convergent

Additional statistically significant differences between housed and unhoused participants occurred regarding reported behavior changes and honesty in the survey responses. Although both housed and unhoused participants noted increased awareness of substance use related to substance use questions, quantitative findings showed that more unhoused participants reported that the study affected their behavior during the week. Qualitatively, we obtained complementary findings in that those who were unhoused more often reported increased substance use, whereas housed participants seemed to mention the awareness of substance use and trended toward a reduction in their use. Finally, convergent findings emerged regarding honesty in the responses. It was clear that housed participants seemed to feel more comfortable answering openly and honestly, perhaps as a result of the newfound freedom associated with transitioning from homelessness to housing [34]. This is contrasted by unhoused participants once again noting paranoia and fear of outing peers on the street.

Discussion

Principal Findings

The results of this mixed methods study illuminate the experiences of housed and unhoused young adults enrolled in an EMA for a 1-week period. Although no statistically significant differences in compliance by housing status were found, statistically significant differences were found regarding

the impact of study participation. Housing status was found to affect young adults' engagement with EMA. Differences in housing status were found regarding the ability to keep their device charged, interference with daily life, stress and anxiety associated with participation, behavior change as a result of EMA, and the ability to respond openly and honestly to prompts.

In terms of compliance, those who had difficulty charging also had lower survey compliance when the phone was charged, and unhoused participants reported greater difficulty in charging. Perhaps the unhoused participants had difficulty charging but made sure to find some way to charge it, possibly because of the importance of the incentive. Similarly, less interference with daily life was associated with greater study compliance in both the daily and EMA surveys, and unhoused individuals noted greater interference. A thought here is the significant correlation between difficulty in charging the device and interference in daily life. Research apps may drain battery life more quickly than other apps and disrupt typical charging patterns based on their usual phone use. If unhoused participants struggled to find a power source and spent much time consumed with finding ways to charge the phone [35] to maintain compliance (ie, get paid), this could interfere greatly with one's day. This could also explain why unhoused participants reported greater stress and anxiety associated with study participation. Again, status was not associated with compliance rates; however, greater stress and anxiety resulting from study participation produced worse compliance rates.

Overall, study compliance was approximately 80%, with no significant detected differences explicitly related to housing status. However, this does not imply that housing status does not need to be considered in the study design. The findings clearly reveal greater impacts, and perhaps burden, for unhoused participants, which was confirmed in the qualitative interviews. Future use of EMA, particularly with unhoused individuals, should consider barriers to using technology in research. Although entirely possible, a study design that relies on the use of technology such as a mobile device should consider issues related to access and how this may create increased burdensomeness. We observed how increased burden could produce additional stress in an already stressful environment, particularly with regard to being tracked. In some cases, the awareness of being tracked has the potential to exacerbate the underlying mental health or substance use issues. Greater efforts to ensure comfort with protocols, including an enhanced focus on confidentiality, could be beneficial, especially because of the histories of marginalization that have led to a deep-seated distrust of systems, including social service systems. Taking more time at the outset of EMA studies to ensure understanding and consent could promote greater trust and decrease the associated stress and anxiety. This could also increase honesty in responses, particularly among unhoused individuals who reported being less honest in their responses compared with those in supportive housing, potentially because of perceived impacts on securing housing or other needed services and resources. Other methods to potentially increase data quality and assurance are to use a lead-in period where participants would get a day or two of practice with the app and EMA questions before recording their responses for analysis.

In testing for reactivity using anxiety and depression symptomatology before and after EMA the week, results show compliance to not be significantly related to the degree of anxiety and depression symptomatology reported. However, a decrease in reported symptoms occurred from before to after the test. This is particularly interesting as some participants, particularly unhoused participants, discussed increased stress and anxiety because of study participation. This indicates that the momentary stress of being prompted did not affect symptomatology and perhaps was fleeting. Although participants may have frequently thought about the bother of the study and became momentarily overwhelmed, it seems to have not been a lasting experience with long-term impacts, despite 22.9% (53/231) of participants feeling that the study affected their behaviors. The fact that compliance was not significantly related to symptomatology (ie, more prompts completed were associated with increases in symptomatology) indicates low reactivity to EMA prompting and participation. More work is needed to effectively examine the overall decreases in symptomatology that may be associated with EMA, as the results cannot definitively be ruled as reactivity, echoing previous literature [36].

The study's implications include support for using intensive longitudinal methods such as EMA with housed and unhoused young adults. Acorda et al [26] explored the impact and acceptability of EMA among 18 youths experiencing homelessness, making recommendations for use with young people who are actively homeless. The results produced by Acorda et al [26] reinforce the findings of this study, most notably the effects of increased self-awareness and the potential for behavior change as a result of EMA. Given the discussion regarding perceived behavioral change and behavioral intentions, EMA also presents possible opportunities for intervention work. It is clear from both their work and the additional support offered from these analyses that EMA is highly acceptable for young adults who have experienced homelessness, both housed and unhoused, with several special considerations, particularly around housing status and confidentiality.

Limitations

Despite the strengths of analyzing this innovative method for vulnerable young adults, there are several limitations that must be acknowledged. The results suggest that some participants did not always answer honestly. We do not have a way of knowing the responses to trust or not using this method with this population. Future work may want to design studies to test the validity of this method with this population, perhaps adapting the study design to include *self-checks* within survey protocols and questions about honesty at the conclusion of each EMA survey. In addition, relying on focus groups as the qualitative methodology could potentially lead to group think more than individual interviews, particularly regarding sensitive topics where group interactions could be detrimental to the discussion. Inquiring about sensitive topics is needed for epidemiological studies; however, a focus solely on what may be perceived as *negative* aspects without the inclusion of a strengths-based perspective lacks an equity lens and was felt by some SPs. Finally, we considered the difference between completion and compliance within the EMA context and ultimately decided to use compliance as the number of prompts completed out of those received based on the study aims. Care should be taken when interpreting findings, specifically if the interest is in distilling information regarding the technological aspects of EMA methods, as this study did not consider technical issues as a focus of analysis, although it is briefly reported.

Conclusions

Although with caveats, this study produced evidence in favor of the use of intensive longitudinal designs with young adults who were formerly and are currently homeless. Intensive longitudinal methods are well suited to capture experiences associated with the chaotic and unstable environments of homelessness, addressing the limitations of cross-sectional and traditional longitudinal designs. However, the findings of this study show that chaotic and unstable environments must be considered at every step of the research process. These findings have implications for research development and design, data collection, and analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Ecological momentary assessment questionnaire.
[[PDF File (Adobe PDF File), 143 KB](#) - [formative_v6i3e33387_app1.pdf](#)]

Multimedia Appendix 2
Daily survey questions.
[[PDF File (Adobe PDF File), 180 KB](#) - [formative_v6i3e33387_app2.pdf](#)]

Multimedia Appendix 3
Focus group guiding questions.
[[PDF File (Adobe PDF File), 56 KB](#) - [formative_v6i3e33387_app3.pdf](#)]

References

1. Smyth JM, Stone AA. Ecological momentary assessment research in behavioral medicine. J Happiness Stud 2003;4(1):35-52. [doi: 10.1023/A:1023657221954]

2. Stone AA, Shiffman S. Ecological momentary assessment (EMA) in behavorial medicine. Ann Behav Med 1994;16(3):199-202. [doi: 10.1093/abm/16.3.199]

3. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. Psychol Assess 2009;21(4):486-497 [FREE Full text] [doi: 10.1037/a0017074] [Medline: 19947783]

4. Trull TJ, Ebner-Priemer U. The role of ambulatory assessment in psychological science. Curr Dir Psychol Sci 2014;23(6):466-470 [FREE Full text] [doi: 10.1177/0963721414550706] [Medline: 25530686]

5. Kuntsche E, Labhart F. Using personal cell phones for ecological momentary assessment: an overview of current developments. Eur Psychol 2013;18(1):3-11. [doi: 10.1027/1016-9040/a000127]

6. Wen CK, Schneider S, Stone AA, Spruijt-Metz D. Compliance with mobile ecological momentary assessment protocols in children and adolescents: a systematic review and meta-analysis. J Med Internet Res 2017;19(4):e132 [FREE Full text] [doi: 10.2196/jmir.6641] [Medline: 28446418]

7. Morton MH, Dworsky A, Matjasko JL, Curry SR, Schlueter D, Chávez R, et al. Prevalence and correlates of youth homelessness in the United States. J Adolesc Health 2018;62(1):14-21 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.10.006] [Medline: 29153445]

8. Gaetz S. Safe streets for whom? Homeless youth, social exclusion, and criminal victimization. Can J Criminol Crim Justice 2004;46(4):423-456. [doi: 10.3138/cjccj.46.4.423]

9. Tyler KA, Hoyt DR, Whitbeck LB, Cauce AM. The effects of a high-risk environment on the sexual victimization of homeless and runaway youth. Violence Vict 2001;16(4):441-455. [Medline: 11506452]

10. Rice E, Rhoades H, Winetrobe H, Sanchez M, Montoya J, Plant A, et al. Sexually explicit cell phone messaging associated with sexual risk among adolescents. Pediatrics 2012;130(4):667-673 [FREE Full text] [doi: 10.1542/peds.2012-0021] [Medline: 22987882]

11. Bender K, Begun S, DePrince A, Haffejee B, Kaufmann S. Utilizing technology for longitudinal communication with homeless youth. Soc Work Health Care 2014;53(9):865-882. [doi: 10.1080/00981389.2014.925532] [Medline: 25321934]

12. Santa Maria D, Padhye N, Yang Y, Gallardo K, Businelle M. Predicting sexual behaviors among homeless young adults: ecological momentary assessment study. JMIR Public Health Surveill 2018;4(2):e39 [FREE Full text] [doi: 10.2196/publichealth.9020] [Medline: 29636318]

13. Zabkiewicz DM, Patterson M, Wright A. A cross-sectional examination of the mental health of homeless mothers: does the relationship between mothering and mental health vary by duration of homelessness? BMJ Open 2014;4(12):e006174 [FREE Full text] [doi: 10.1136/bmjopen-2014-006174] [Medline: 25492272]

14. Jones AA, Gicas KM, Seyedin S, Willi TS, Leonova O, Vila-Rodriguez F, et al. Associations of substance use, psychosis, and mortality among people living in precarious housing or homelessness: a longitudinal, community-based study in Vancouver, Canada. PLoS Med 2020;17(7):e1003172 [FREE Full text] [doi: 10.1371/journal.pmed.1003172] [Medline: 32628679]

15. Santa Maria D, Padhye N, Yang Y, Gallardo K, Santos GM, Jung J, et al. Drug use patterns and predictors among homeless youth: results of an ecological momentary assessment. Am J Drug Alcohol Abuse 2018;44(5):551-560. [doi: 10.1080/00952990.2017.1407328] [Medline: 29286835]

16. Tyler KA, Olson K, Ray CM. Understanding the link between victimization and alcohol use among homeless youth using ecological momentary assessment. Socius 2018;4:2378023118779832 [FREE Full text] [doi: 10.1177/2378023118779832] [Medline: 31179380]

17. Tyler KA, Olson K. Examining the feasibility of ecological momentary assessment using short message service surveying with homeless youth: lessons learned. Field methods 2018;30(2):91-104 [FREE Full text] [doi: 10.1177/1525822x18762111] [Medline: 32863777]

18. Nelson RO. Assessment and therapeutic functions of self-monitoring. Prog Behav Modif 1977;5:263-308. [doi: 10.1016/B978-0-12-535605-3.50012-1]

19. van Ballegooijen W, Ruwaard J, Karyotaki E, Ebert DD, Smit JH, Riper H. Reactivity to smartphone-based ecological momentary assessment of depressive symptoms (MoodMonitor): protocol of a randomised controlled trial. BMC Psychiatry 2016;16(1):359 [FREE Full text] [doi: 10.1186/s12888-016-1065-5] [Medline: 27769201]

20. Husky M, Olié E, Guillaume S, Genty C, Swendsen J, Courtet P. Feasibility and validity of ecological momentary assessment in the investigation of suicide risk. Psychiatry Res 2014;220(1-2):564-570. [doi: 10.1016/j.psychres.2014.08.019] [Medline: 25155939]

21. Weatherson K, Yun L, Wunderlich K, Puterman E, Faulkner G. Application of an ecological momentary assessment protocol in a workplace intervention: assessing compliance, criterion validity, and reactivity. J Phys Act Health 2019;16(11):985-992. [doi: 10.1123/jpah.2019-0152] [Medline: 31541068]

22. Carr DJ, Adia AC, Wray TB, Celio MA, Pérez AE, Monti PM. Using the internet to access key populations in ecological momentary assessment research: comparing adherence, reactivity, and erratic responding across those enrolled remotely versus in-person. Psychol Assess 2020;32(8):768-779 [FREE Full text] [doi: 10.1037/pas0000847] [Medline: 32437190]

23. Munsch S, Meyer AH, Milenkovic N, Schlup B, Margraf J, Wilhelm FH. Ecological momentary assessment to evaluate cognitive-behavioral treatment for binge eating disorder. Int J Eat Disord 2009;42(7):648-657. [doi: 10.1002/eat.20657] [Medline: 19197978]

24. Hufford MR, Shields AL, Shiffman S, Paty J, Balabanis M. Reactivity to ecological momentary assessment: an example using undergraduate problem drinkers. Psychol Addict Behav 2002;16(3):205-211. [Medline: 12236455]

25. Heron KE, Smyth JM. Is intensive measurement of body image reactive? A two-study evaluation using ecological momentary assessment suggests not. Body Image 2013;10(1):35-44 [FREE Full text] [doi: 10.1016/j.bodyim.2012.08.006] [Medline: 22999225]

26. Acorda D, Businelle M, Santa Maria D. Perceived impacts, acceptability, and recommendations for ecological momentary assessment among youth experiencing homelessness: qualitative study. JMIR Form Res 2021;5(4):e21638 [FREE Full text] [doi: 10.2196/21638] [Medline: 33821805]

27. Wolfe SM, Toro PA, McCaskill PA. A comparison of homeless and matched housed adolescents on family environment variables. J Res Adolesc 1999;9(1):53-66. [doi: 10.1207/s15327795jra0901_3]

28. Semborski S, Henwood B, Rhoades H, Mason T, Wenzel S, Rice E. Construct, concurrent, and real-world predictive validity of the Difficulties in Emotion Regulation (DERS-18) among young adults with history of homelessness. Psychol Assess 2021;33(5):385-394. [doi: 10.1037/pas0000991] [Medline: 33630634]

29. Gewirtz O'Brien JR, Edinburgh LD, Barnes AJ, McRee AL. Mental health outcomes among homeless, runaway, and stably housed youth. Pediatrics 2020;145(4):e20192674. [doi: 10.1542/peds.2019-2674] [Medline: 32152134]

30. Kemp PA, Neale J, Robertson M. Homelessness among problem drug users: prevalence, risk factors and trigger events. Health Soc Care Community 2006;14(4):319-328. [doi: 10.1111/j.1365-2524.2006.00624.x] [Medline: 16787483]

31. Henwood BF, Redline B, Dzubur E, Madden DR, Rhoades H, Dunton GF, et al. Investigating health risk environments in housing programs for young adults: protocol for a geographically explicit ecological momentary assessment study. JMIR Res Protoc 2019;8(1):e12112 [FREE Full text] [doi: 10.2196/12112] [Medline: 30632969]

32. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

33. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

34. Henwood BF, Derejko KS, Couture J, Padgett DK. Maslow and mental health recovery: a comparative study of homeless programs for adults with serious mental illness. Adm Policy Ment Health 2015;42(2):220-228 [FREE Full text] [doi: 10.1007/s10488-014-0542-8] [Medline: 24518968]

35. Galperin H, Bar F, Nguyen H. The power divide: mobile communication in Los Angeles' skid row. Mob Media Commun 2021;9(1):30-50. [doi: 10.1177/2050157920932608]

36. Rowan PJ, Cofta-Woerpel L, Mazas CA, Vidrine JI, Reitzel LR, Cinciripini PM, et al. Evaluating reactivity to ecological momentary assessment during smoking cessation. Exp Clin Psychopharmacol 2007;15(4):382-389. [doi: 10.1037/1064-1297.15.4.382] [Medline: 17696685]

Abbreviations

EMA: ecological momentary assessment
SP: study participant

XSL•FO
RenderX

<u>Original Paper</u>

Preliminary Exploration of Main Elements for Systematic Classification Development: Case Study of Patient Safety Incidents

Riikka Vuokko[1*], PhD; Anne Vakkuri[2*], MD, PhD; Sari Palojoki[1*], PhD

[1]Department of Steering of Health Care and Social Welfare, Ministry of Social Affairs and Health, Helsinki, Finland

[2]Department of Anesthesiology, Intensive Care and Pain Medicine, Peijas Hospital, Helsinki University Hospital, Vantaa, Finland

[*]all authors contributed equally

Corresponding Author:
Riikka Vuokko, PhD
Department of Steering of Health Care and Social Welfare
Ministry of Social Affairs and Health
P.O. Box 33
Helsinki, 00023
Finland
Phone: 358 29516001
Email: riikka.vuokko@gov.fi

Abstract

Background: Currently, there is no holistic theoretical approach available for guiding classification development. On the basis of our recent classification development research in the area of patient safety in health information technology, this focus area would benefit from a more systematic approach. Although some valuable theoretical and methodological approaches have been presented, classification development literature typically is limited to methodological development in a specific domain or is practically oriented.

Objective: The main purposes of this study are to fill the methodological gap in classification development research by exploring possible elements of systematic development based on previous literature and to promote sustainable and well-grounded classification outcomes by identifying a set of recommended elements. Specifically, the aim is to answer the following question: what are the main elements for systematic classification development based on research evidence and our use case?

Methods: This study applied a qualitative research approach. On the basis of previous literature, preliminary elements for classification development were specified, as follows: defining a concept model, documenting the development process, incorporating multidisciplinary expertise, validating results, and maintaining the classification. The elements were compiled as guiding principles for the research process and tested in the case of patient safety incidents (n=501).

Results: The results illustrate classification development based on the chosen elements, with 4 examples of technology-induced errors. Examples from the use case regard usability, system downtime, clinical workflow, and medication section problems. The study results confirm and thus suggest that a more comprehensive and theory-based systematic approach promotes well-grounded classification work by enhancing transparency and possibilities for assessing the development process.

Conclusions: We recommend further testing the preliminary main elements presented in this study. The research presented herein could serve as a basis for future work. Our recently developed classification and the use case presented here serve as examples. Data retrieved from, for example, other type of electronic health records and use contexts could refine and validate the suggested methodological approach.

Introduction

Background

Classifications are applied for various purposes in clinical contexts, including patient documentation and more specific domains, such as the patient safety incident reporting of technology-induced errors. However, the availability of methodological frameworks and research evidence is limited for developing and maintaining clinical classifications [1-3]. Typically, classification development has practical goals in a specific clinical setting or documentation context. Accordingly, the results of practice-driven classification work are not necessarily transferable to other organizations, clinical domains, or purposes [4,5].

According to literature, severe challenges have arisen from the emergence of safety issues involving newly innovated and existing implemented health information technologies (ITs) [6-8]. Regarding patient safety in health IT, previous classification development-related research on technology-induced errors is available [3,9,10]. Magrabi et al [3] identified categories for populating the classification of IT problems with the aim of providing a clinical method for classifying computer-related patient safety incidents. This classification was then tested with patient safety incident data in a setting with 100% coverage of electronic health records (EHRs). The results indicated a need for classification development from the perspective of a sociotechnical approach [1]. Moreover, a recent study summarized that organizations today do not have rigorous, real-time approaches for routinely assessing the safety of EHRs or for identifying safety hazards [11]. This constitutes the underlying rationale for why we have recently developed and validated a classification of technology-induced error types from a social-technical perspective [12]. A guiding assumption in our research was that despite previous research introducing relevant themes in this area, further research is needed to capture the fast-developing clinical working environment. Health IT and EHRs advance in ways that require the identification and classification of new types of phenomena. Classifications evolve as new evidence arises [13].

On the basis of our experience when researching classification development, this focus area benefits from a more systematic approach. Although valuable theoretical and methodological approaches have been presented, some of the literature concentrates on a specific part of the classification development, such as validation within a specific domain [14] or maintenance and implementation of international classifications [15]. No holistic approach is provided to guide classification development. In addition, although some valuable research on patient safety incidents and respective classifications has been conducted, there is relatively scarce documentation of detailed classification development that could guide future development. It is suggested [16] that publishing methodological and theoretical approaches applied in classification development cases increases the validity and quality of development outcomes. This contributes in building a systematic approach to classification development. Currently, for example, some

practical guidelines and web-based resources are available regarding large-scale international examples, such as the World Health Organization's International Classification of Diseases and Systematized Nomenclature of Medicine–Clinical Terms [17,18]. However, the gap between practical guidelines and theory-based systematic classification development presents a challenge when the aim is to develop most sustainable, well-grounded classifications, which can also be assessed in a transparent way against a set of common development principles [16,19].

Objectives

To benefit future classification work, in this paper, we present preliminary results on methodological considerations based on previous research evidence and our case research on classifying EHR-related patient safety incidents [12]. Our preliminary results are intended to inform future research on classification development research protocols with different use cases. We characterize typical elements of classification development and suggest a methodological approach [20] to achieve more systematic classification development after testing it with other use cases. Therefore, our main research question is as follows: what are the main elements for systematic classification development based on research evidence and our use case?

Methods

Evidence From Previous Research Informing Our Study Design

Sociotechnical theory has been a basis for the development of health IT-related models [21]. Pioneering theorists of the sociotechnical approach emphasized that technical and social systems should be optimized simultaneously and that organizations should comprise a relationship between nonhuman and human systems [22,23]. More specifically, the theoretical background for developing our classification was the Health IT Safety measurement framework proposed by Sittig and Singh [9,24]. The starting point of the Health IT Safety framework is that safety incidents must be understood within the full context of the *sociotechnical work system*. This refers to the interacting technical (eg, hardware and software) and nontechnical (eg, clinical workflow, people, and the physical environment) variables that affect health IT-related patient safety. The framework responds to the fundamental conceptual and methodological gaps related to both defining and measuring health IT-related patient safety. The aim is to provide a conceptual foundation for health IT-related patient safety measurement, monitoring, and improvement [10]. The following three domains were created to describe the range of risks and opportunities of health IT to influence patient safety: Safe Health IT, Safe Use of Health IT, and Using Health IT to Improve Safety. The latter includes using technology to identify and monitor patient safety incidents, risks, and hazards and to intervene before harm occurs [24].

EHR-related classifications for the purpose of incident reporting have been developed for specific clinical settings and problem areas. For example, the French Nuclear Safety Authority scale for classifying incidents was applied in oncology to inform the

design and potential utility of an incident reporting system. All incidents during the research period were reviewed and graded according to potential severity by the consensus of a committee, including physicians and physicists [25]. However, the aim of the research was not to provide a detailed description of the development process of IT-related error-type categories from a methodological classification development perspective. Another example of classification-related development work in a specific IT-related error area is the development of a usability-related error ontology by Elkin et al [26]. Here, initial semantics for usability error types were derived from a literature review and an expert opinion. Then, a participatory design method was used to obtain input and feedback from multi-professional stakeholders. According to Elkin et al [26], with use and experience, the ontology would grow and evolve toward more standard and interoperable reporting. Intrinsically, ontologies have become important resources, because they can be applied to integrated EHR infrastructures to improve possibilities of data acquisition and storage, standardization, interoperation, data analysis for clinical research, and routine clinical documentation [27,28].

The classification developed by Magrabi et al [3] is a rare example of EHR-related classification development research that is not restricted to a specific EHR problem area. It was developed by using safety incidents and was further refined by analyzing data from incident reports submitted to a regulatory database. The methodology of classification development was based on the free-text descriptions of a quarter of the incidents retrieved (n=123), which were used to identify *natural categories* for the classification. A simple classification of the reported problems related to computer use was developed. The incidents were classified using the classification. An interrater reliability analysis was performed using the κ statistic to determine consistency among researchers. The classification developed by Magrabi et al [3] underwent testing in the United Kingdom after the preliminary development phases. In subsequent research, a limitation of the classification was documented: it was not possible to demonstrate the clinical relevance of all incidents by using this classification [29]. However, the study focused on case data–driven development of the classification, not on describing the underlying methodological aspects of classification theories.

In a recent study [12], a classification for patient safety incident reporting associated with the use of a mature EHR was developed, which was validated using a data set of 501 patient safety incidents. Here, a mature EHR is defined according to the Electronic Medical Record Adoption Model, which was developed by Healthcare Information and Management Systems Society Analytics. This universally recognized maturity model is an 8-stage model that reflects hospitals' electronic medical record capabilities, ranging from a completely paper-based environment (stage 0) to a highly advanced paperless and digital patient record environment (stage 7). Regarding these data, the maturity level of the EHR system is 6 to 7 [12]. The classification development was based on research into commonly recognized error types. Further, a multi-professional research team used iterative tests on consensus building to develop a classification and preliminary descriptions of the classes. The final classification was validated using incident report data to evaluate its characteristics and applicability for purposes of patient safety incident reporting. The development focused on applying the theoretical aspects of classification development, for example, by defining concepts and the exclusiveness of categories and forming descriptions for all categories with the quality and usability of the resulting classification as a guiding principle. This classification development and validation research was used as a use case in this research to strengthen the fragmented methodological support for this type of research.

Methodological Background and Study Design

This study applied a qualitative research approach that is applicable for identifying, characterizing, and interpreting a phenomenon. Qualitative methodology relies on data and their interpretation according to respective conceptualization [20,30]. To increase the reliability of qualitative research, applying a varied methodology is suggested [20,30-32]. For our use case [12], this involved retrieving clinical patient safety incident reporting data and conceptualizing technology-induced errors based on previous research, selecting relevant reports, analyzing incident data to define categories and hierarchies for the emerging classification, and applying a multidisciplinary panel within our research team to reach agreement over ongoing classification development (Multimedia Appendix 1). We have already documented the study design for data analysis and validation [12], and in this paper, our study design concentrates on specifically capturing classification development characteristics (Textbox 1) and related observations from the use case of our classification development. By creating a holistic methodological approach for classification development, there is potential to expand use to different kind of contexts and use cases.

Textbox 1. Main elements of classification development based on previous research.

1. A concept model: a concept model supports the data analysis required in classification development. A model includes, for example, the name and version of the classification, names of each category and their descriptions, possible translations, types of relations among categories, and mapping data regarding terminology harmonization [4,33,34].

2. Documentation of classification development: documenting the elements and progress of classification development in sufficient detail adds to the understanding of concept analysis and building relationships among categories, with an emphasis on clear separation among categories [4,16,33].

3. Multidisciplinary expertise: for example, clinical, terminological, informatics, and technical expertise used in the analysis and development process ensures relevant and usable outcomes [4,12,16,26,27].

4. Validation of results: development result validation based on study design and expertise or, for example, the translation of concepts and categories to ensure that the chosen concepts remain unchanged and the clinical meaning prevails. This supports interoperability and is especially relevant for small languages, such as Finnish [12,20,35,36].

5. Classification maintenance: after developing a classification, the formal governance model or informal maintenance process should include feedback from the clinicians (ie, the users) [34,37,38].

Classification development research is considered more practical than theoretically oriented, which results in development outcomes that might not necessarily be scalable to other research or use contexts [16,39,40]. Unsystematic and partial descriptions of classification development steps render it difficult to understand the underlying theoretical grounds for created structures and characteristics [33]. Applying a qualitative approach allows us to use somewhat limited research literature both as an analytical tool and to provide a source of concepts, theories, and hypotheses [20,31]. For qualitative research, a systematic review of literature is not typically required, although familiarity with relevant literature may increase sensitivity to subtle nuances of data [20,32]. In line with the qualitative approach [20], we identified key elements of classification development from the previous literature. The main literature consists of 12 research papers, and the results are summarized in Textbox 1. With this use case, we applied the elements found in the literature by researchers who described such elements and stages of classification development. Therefore, before embarking on classification development, typical elements based on the previous literature were compiled as guiding principles in the research process. These guiding principles were used and tested in practice during our recent study [12] (Textbox 1). However, these elements were not described in-depth in our publication.

According to the available literature, the classification development process begins with a careful, in-depth concept analysis to support the need for structured and controlled data representation (cf the study by Watson et al [16]). This is for documentation purposes [27] and for increasing the usability of data by ensuring improved data quality and comparability [41]. Reconciling clinical needs for documentation requires content analysis and mapping, which can be costly in both financial and temporal respects [4]. An important observation for classification development is that incomplete or overlapping conceptualizing, naming, and descriptions of categories and their relations within a classification may challenge resulting data quality owing to heterogeneity and indistinctness of concepts and terms used [4,33]. We have defined the concepts used in our research in the previous paper [12]. Our core concept regarding the phenomenon of the use case is a technology-induced error that results from the design and

development of technology (Multimedia Appendix 2), the implementation and customization of technology, and the interplay between the operation of a technology and the new work processes that arise from the use of technology [42,43]. Regarding the classification category and subcategory building, our process was both qualitative and iterative [20]. The category building consisted of systematic analysis and labeling of varied phenomena illustrated in our data [20]. By examining differences and similarities in our data, we divided the data first into categories, iteratively refined distinctive differentiation between categories and within a category, and continued to define subcategories with an in-depth analysis.

From a methodological perspective, regarding the aim of improving systematic classification development, the potential strengths and weaknesses encountered during the process should be documented in research. Cornet et al [4] stated that content shortcomings encountered during classification development, for example, concept coverage and gaps, can be solved relatively easily. However, some formalism issues are more difficult to resolve. These include, for example, how relations among the categories are arranged, the overall structure or concept model, and how the classification is applied by the clinical users [27,41]. The United Nations suggests the following similar requirements in their practical guideline for classification development, for example, a consistent conceptual basis, a flat or hierarchic structure, categories that are mutually exclusive and exhaustive, and definitions that are clear and unambiguous and define the content of each category [44]. Furthermore, classifications should be relevant to users and sufficiently robust to last for a period of intended use. They should also provide comparability over time and among collections and provide guidelines for the coding and output of any data collected [36,38].

To summarize, as described in our study design, these elements form the basis for testing them in our use case.

Results

Overview

Our results provide examples of selected categories of our use case analysis to illustrate the development of technology-induced problem classification. During the analysis [12], we reviewed

all patient safety incidents (N=1486) to identify technology-induced ones and continued to categorize the remaining incidents (n=501) with the concepts based on previous research. We added both main categories and subcategories to develop the initial content toward the context of mature EHRs according to evidence from the patient safety incident data. This required close cooperation between clinical and informatics experts within the research team to ensure that the categories captured identified phenomena correctly and were relevant for clinicians.

We provide the following four examples from our use case [12]: a description of the reasoning behind 2 categories, which were expanded to increase the potential accuracy of patient safety incident reporting, and 2 new categories, which were added to reduce ambiguities and any duplication of the original categories.

Category of Usability Issues

The first example is illustrated by the category of usability issues (73/501, 17.1%). To start with, this category had no description, and the initial subcategories included features that were more closely related to other categories, such as documentation problems (60/501, 14.1%; Multimedia Appendix 2). With the help of clinical experts, we defined usability problems as situations in which the used EHR is difficult to use. For example, this complicates finding the required patient data or does not guide the clinician as expected, meaning that the system does not support the clinician's work processes as expected according to care guidelines. In contrast, the documentation problems were defined to include a lack of data structure, errors, and ambiguities when entering data. After the analysis, we added 4 subcategories to usability issues, resulting in a 2-level hierarchy. Two of these subcategories (problems with decision support, n=2, and printing problems, n=11) were easily defined. For the other two subcategories (problems and deficiencies related to alarms, n=29, and problems with finding data, n=30), we added additional third-level subcategories. In addition, we added missing, incorrect, or difficult-to-interpret alarm and alarm fatigue to the first subcategory. For the second subcategory, we added that the information is difficult (illogical) for the user to perceive and the information is laborious to find and has to be dug up. These third-level subcategories were easily identified by the clinicians in our research team and are documented in previous research. Although they concern separate usability issues, only future use of the classification would reveal if the third-level subcategories are suitable for reporting. In an ideal situation, third-level subcategories could guide the reporting clinician to identify a usability issue more easily, based on the category-specific name and description. However, a deeper hierarchy of categories and additional subcategories could increase the reporting burden in fast-paced clinical work. To determine the most suitable level of accuracy for reporting, the pilot use of the developed classification would be required.

Category of EHR Downtime

Regarding the category of EHR downtime (8/501, 1.9%), the starting point for classification development was a flat hierarchy, that is, a single-level main category, after which our development progressed iteratively. This category was identified

as a relatively well-defined entity with no connection to other categories in the classification. Furthermore, it was known from the literature that domain coverage (including only planned and unplanned downtimes based on previous research at this stage) was insufficient to provide classification benefits. However, we sought a relatively simple 3-level hierarchical structure to facilitate clinical use. In addition to the literature, available real-world data in the form of patient safety incident reports guided more precise identification of new subcategories, including the completely new phenomenon of *the problem of logging on to a single application partly or entirely* (n=5). At this point, we divided the subcategory of the system logging problem in relation to the whole or part of the system in use to capture issues with a high-maturity EHR more effectively. The harmonization of concepts for known phenomena based on previous literature was relatively straightforward. However, the concepts for optimizing the new range of categories for clinical users required more reconciliation. Our aim is that categories would be mutually exclusive and exhaustive. During classification work with our data, we noticed that not all incidents fitted with existing subcategories. At this point, attention was given to content according to the feedback of clinicians as users. Without incorporating multidisciplinary knowledge, it would have been difficult to create and define content clearly and comprehensively, especially for the category *data entry during and after a downtime*.

Category of Clinical Workflow Problems

During classification development, we added clinical workflow problems (33/501, 7.7%) as a new main category for two main reasons: the original classification addresses this phenomenon ambiguously, and mature EHRs implement clinical procedures according to defined workflow descriptions and guidelines. In other words, although building a workflow into EHRs is a well-established practice, the original classification left the issue either unidentified or partially identified. On the basis of the use case and the views of clinical experts, workflow problems can disrupt work procedure continuity. For example, although fundamental process steps such as a patient transfer or discharge should be carried out, the EHR renders it impossible to complete the procedure or does not otherwise support the clinician's work as expected according to hospital guidelines. Moreover, owing to the lack of an integrated, functional, and logical workflow, the system may attract ways of clinical use that are not aligned with the guidelines when the expected EHR functionalities cannot be obtained. This can result in deviant work processes and workarounds. On the basis of our case data, the phenomena of workflow problems could be accurately identified and tentatively described. However, this could still benefit from insights from other use cases and data to confirm the correctness of our interpretation based on these specific data. Moreover, identifying workflow problems as a new main category can reduce the corruption of reported data caused by the weak distinctiveness of the original classification and its categories.

Category of Medication Section Problems

We noticed that the original classification did not cover incidents related to the medication management section adequately, especially given that many of these cases were documented in

the research data. Thus, a new main class of medication section problems (89/501, 20.8%) was added. These problems can result in situations where prescription and patient record information is not stored as intended, or where apparent changes occur because of an unidentified system-related reason. This is often a problem that hinders the management of overall medication. With this particular category based on its appearance in our research data, we wish to illustrate that there are some preconditions for classification development. First, the problems in the medication section appeared more difficult to delimit and describe compared with workflow problems. However, we recognized that these problems are closely related to the clinical workflow category features (eg, functionality of closed-loop medicine administration) and documentation category. Regarding medication section incidents, our research team concluded that there were still complex unresolved issues in the implementation of the new EHR from the implementation perspective, where immediate modification and mitigation appears to be necessary. Thus, it is likely that some of the system-specific problems will be solved. Moreover, when identifying real root causes to develop new subcategories, a more detailed examination would have been required, for example, through root cause analysis. In this case, it was too early to develop subcategories based on the research data; hence, the problem type could be identified but not unambiguously described. For this category, we found that a major functional change period, that is, system implementation in progress, may be a suboptimal time for developing a classification.

To conclude the insights from the use case, several iterations of category names and descriptions were required to ensure shared and sufficient understanding of the reasoning behind a specific category. Further, we used research literature to determine how the same or similar categories had been conceptualized. During classification development, clinician insights provided clinical understanding of a phenomenon through a clinical lens. However, informatics and IT expertise were required to identify and analyze system-specific starting points and boundaries of a specific phenomenon. Overall, our classification development was conducted through a sociotechnical lens to ensure comparable reasoning of both human and nonhuman factors that attribute to a specific technology-induced patient safety incident. However, when new categories or subcategories were constructed, the clinical perspective of relevant and usable classification content development remained a prevailing guideline. This was particularly true regarding clinical reporting and the optimal versus redundant level of accuracy for reporting based on the available categories and time available for reporting in hectic clinical work. To summarize, essential key elements, such as content analysis, category building, the definition of hierarchy among categories, and classification validation, provide requirements for systematic classification development.

Discussion

Principal Findings

On the basis of the study design, a set of guiding principles and preliminary elements for classification development were identified and applied in our use case. Next, we gathered evidence of the process for classification development and captured common elements based on previous research and our own experience of classification development. Finally, we processed these elements and explored examples from our recent research [12] in the *Results* section to illustrate the details of both the classifying process and elements to build sustainable and well-defined classifications in the future.

Our classification development and validation research [12] was used as a use case in this research to strengthen the fragmented methodological support for this type of research. In practice, we revealed incident-related gaps of conceptualization within the area of research. During classification development, filling these gaps meant analyzing various technology-induced patient safety incidents within our multidisciplinary research team. The incidents were analyzed with a sociotechnical lens to avoid bias when building the categories. Clinician insights into the incidents and the rationality of the clinical working environment played significant roles when identifying the categories.

Owing to the scarcity of research evidence, challenges when conducting this kind of research are recognized [4,12]. Accordingly, our main purpose is to fill a methodological gap in classification development research to promote well-grounded classification development. Although we ground these recommendations on the use case, the available research literature indicates similar challenges in other classification development cases. First, a concept model informing the classification development promotes systematic category building with structural formalism. In a way, such a model provides fundamental requirements for the development of each category. Second, in addition to conceptualizing category building, documenting agreements on other main elements can constitute guiding principles for the whole classification development. Furthermore, documenting common agreements can also support quality control and ensure the understanding of common goals for the development. Third, ensuring multidisciplinary expertise is crucial for guaranteeing that the intended clinical meaning prevails and for facilitating the validation of emerging classes. Moreover, regarding the future use of a classification under development, multidisciplinary expertise can be used to define the required level of category robustness and to review sufficient levels of clinical domain coverage and comprehensiveness. By reviewing clinical relevance and documenting practices during classification development, the development outcome might be feasible for clinical use. Fourth, terminological harmonization or systematic translation of concepts and terms representing the categories can be used to ensure that the clinical meaning remains unchanged during the development and that the chosen concepts remain relevant in the specific domain or clinical context. Fifth, planning ahead is a requirement of classification maintenance, especially when the clinical context in which the classification is used is constantly changing and evolving. In this regard, planning for user feedback or other routes for receiving and processing development needs emerging from the clinical context and users in the future can provide a means of classification maintenance.

Limitations

Any method of qualitative or quantitative analysis is not a purely technical process, as it is inevitably influenced by the scientific background and experience of researchers. Accordingly, critical reflection throughout the research process is of considerable importance. The multidisciplinary research team with prior training and experience in qualitative analysis worked closely to follow up on research progress analytically and critically to ensure the quality of the conduct of the study. However, there are many known weaknesses associated with the qualitative research approach in particular [20,30-32]. In this study, the approach was iterative, the analysis was interpretative, and the results were consequently descriptive, which is typical of this research approach. Therefore, the results should be considered from the outset of these facts. The most difficult methodological question arose from the situation where the researchers determined that a systematic review of the literature in this area does not necessarily produce such results that would benefit the achievement of the research goals. Moreover, the researchers have been working on governmental and international classification development for many years, which contributed to the decision to proceed in a way that deviated from the original research plan, which relied on the possibilities provided by the qualitative research literature. Furthermore, the literature was studied and adopted cumulatively.

The theoretical framework for our classification development was the sociotechnical model [9,10]. It should be noted that classification development can be based on many starting points and can occur within many themes. As a result, we cannot conclude from this use case alone that our approach would be applicable to all subject areas. This is why further research is still required, in which our observations could be tested with new data.

Lessons Learned and Implications for Future Work

There are many possible aspects for future classification development work and related research based on our study results. Based on our observations, for the continuum of classification development in each context, we suggest that sufficient multi-professional analysis and review should be part of classification development and maintenance. Overall, the classification development should place more emphasis on terminological and clinical subject area expertise.

Moreover, classifications can be regarded as a representation framework among evolving practices, meaning they must be followed by new knowledge. Furthermore, more case or pilot studies with real-world contexts are required. Our use case and developed classification serve as an example here: patient safety incident data retrieved from other source EHRs and use contexts could strengthen the methodological approach or it might contribute to further development of our classification based on user feedback.

As we became acquainted with the classification literature, preliminary observations were made about promising future possibilities, specifically of ontologies. Although we excluded the topic from our own research, we highlight the potential for future research. Hancock [41] described a study that demonstrates the demand for instant data and information is enabled through innovative and newer ways of classifying information. From a system perspective, much of the development process can now be automated, content can be contributed and approved on the web, and computer programs are sufficiently advanced to consider more human-thinking methodologies. Although it will take time to replace traditional approaches to classification development and theory with innovative, technological solutions, we suggest re-examining the classification development process we have described, starting with the concept definition phase. It would be particularly interesting to determine whether machine learning can be used in this context, especially if the aim is to analyze extensive data sets.

Conclusions

Although classifications remain significant tools for clinical documentation and for producing clinical data in various clinical domains, limited research literature is available that illustrates classification development systematically or from a methodological perspective based on previous research. Owing to the role of classifications in data production, theoretical and systematic reviews could also contribute to the transparent development of a future health care knowledge base. Thus, we recommend the main elements based on this study for systematic classification development. Furthermore, the research presented herein could serve as a basis for future work. To conclude, there is a need for the scientific assessment of whether classifications in different domain areas can be developed from theoretical and systematic perspectives.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of the study design.
[DOCX File , 38 KB - formative_v6i3e35474_app1.docx]

Multimedia Appendix 2

XSL•FO
RenderX

Concept model of the use case specific classification development.
[DOCX File , 448 KB - formative_v6i3e35474_app2.docx]

References

1. Palojoki S, Mäkelä M, Lehtonen L, Saranto K. An analysis of electronic health record-related patient safety incidents. Health Informatics J 2017;23(2):134-145 [FREE Full text] [doi: 10.1177/1460458216631072] [Medline: 26951568]

2. Meeks DW, Smith MW, Taylor L, Sittig DF, Scott JM, Singh H. An analysis of electronic health record-related patient safety concerns. J Am Med Inform Assoc 2014;21(6):1053-1059 [FREE Full text] [doi: 10.1136/amiajnl-2013-002578] [Medline: 24951796]

3. Magrabi F, Ong MS, Runciman W, Coiera E. An analysis of computer-related patient safety incidents to inform the development of a classification. J Am Med Inform Assoc 2010;17(6):663-670 [FREE Full text] [doi: 10.1136/jamia.2009.002444] [Medline: 20962128]

4. Cornet R, de Keizer NF, Abu-Hanna A. A framework for characterizing terminological systems. Methods Inf Med 2006;45(3):253-266. [Medline: 16685333]

5. Palojoki S, Vuokko R, Vakkuri A, Saranto K. Electronic health record system-related patient safety incidents - how to classify them? Stud Health Technol Inform 2020;275:157-161. [doi: 10.3233/SHTI200714] [Medline: 33227760]

6. Borycki E. Quality and safety in eHealth: the need to build the evidence base. J Med Internet Res 2019;21(12):e16689 [FREE Full text] [doi: 10.2196/16689] [Medline: 31855183]

7. Carayon P, Du S, Brown R, Cartmill R, Johnson M, Wetterneck TB. EHR-related medication errors in two ICUs. J Healthc Risk Manag 2017;36(3):6-15 [FREE Full text] [doi: 10.1002/jhrm.21259] [Medline: 28099789]

8. Graber ML, Byrne C, Johnston D. The impact of electronic health records on diagnosis. Diagnosis (Berl) 2017;4(4):211-223 [FREE Full text] [doi: 10.1515/dx-2017-0012] [Medline: 29536944]

9. Singh H, Sittig DF. A sociotechnical framework for safety-related electronic health record research reporting: the SAFER reporting framework. Ann Intern Med 2020;172(11 Suppl):S92-100 [FREE Full text] [doi: 10.7326/M19-0879] [Medline: 32479184]

10. Singh H, Sittig DF. Measuring and improving patient safety through health information technology: the Health IT Safety Framework. BMJ Qual Saf 2016;25(4):226-232 [FREE Full text] [doi: 10.1136/bmjqs-2015-004486] [Medline: 26369894]

11. Sittig DF, Wright A, Coiera E, Magrabi F, Ratwani R, Bates DW, et al. Current challenges in health information technology-related patient safety. Health Informatics J 2020;26(1):181-189 [FREE Full text] [doi: 10.1177/1460458218814893] [Medline: 30537881]

12. Palojoki S, Saranto K, Reponen E, Skants N, Vakkuri A, Vuokko R. Classification of electronic health record-related patient safety incidents: development and validation study. JMIR Med Inform 2021;9(8):e30470 [FREE Full text] [doi: 10.2196/30470] [Medline: 34245558]

13. Cree IA, Indave BI. Commentary: cancer research quality and tumour classification. Tumour Biol 2020;42(2):1010428320907544 [FREE Full text] [doi: 10.1177/1010428320907544] [Medline: 32098581]

14. Van Tiggelen H, LeBlanc K, Campbell K, Woo K, Baranoski S, Chang YY, et al. Standardizing the classification of skin tears: validity and reliability testing of the International Skin Tear Advisory Panel Classification System in 44 countries. Br J Dermatol 2020;183(1):146-154 [FREE Full text] [doi: 10.1111/bjd.18604] [Medline: 31605618]

15. Harrison JE, Weber S, Jakob R, Chute CG. ICD-11: an international classification of diseases for the twenty-first century. BMC Med Inform Decis Mak 2021;21(Suppl 6):206 [FREE Full text] [doi: 10.1186/s12911-021-01534-6] [Medline: 34753471]

16. Watson J, Nicholson BD, Hamilton W, Price S. Identifying clinical features in primary care electronic health record studies: methods for codelist development. BMJ Open 2017;7(11):e019637 [FREE Full text] [doi: 10.1136/bmjopen-2017-019637] [Medline: 29170293]

17. Maintenance platform for the WHO family of international classifications (WHOFIC). World Health Organization. URL: https://icd.who.int/dev11 [accessed 2021-08-20]

18. SNOMED CT editorial guide. SNOMED International. URL: https://confluence.ihtsdotools.org/display/DOCEG [accessed 2021-11-08]

19. Audigé L, Bhandari M, Hanson B, Kellam J. A concept for the validation of fracture classifications. J Orthop Trauma 2005;19(6):401-406. [doi: 10.1097/01.bot.0000155310.04886.37] [Medline: 16003200]

20. Strauss A, Corbin J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. 2nd Edition. Thousand Oaks, CA: Sage Publications; 1998.

21. Borycki EM, Kushniruk AW. Towards an integrative cognitive-socio-technical approach in health informatics: analyzing technology-induced error involving health information systems to improve patient safety. Open Med Inform J 2010;4:181-187 [FREE Full text] [doi: 10.2174/1874431101004010181] [Medline: 21594010]

22. Emery FE, Trist EL. The causal texture of organizational environments. Hum Relat 1965;18(1):21-32. [doi: 10.1177/001872676501800103]

23. Trist EL. The Evolution of Socio-Technical Systems: A Conceptual Framework and An Action Research Program. Ontario, Canada: Ontario Ministry of Labour, Ontario Quality of Working Life Centre; 1981.

24. Sittig DF, Singh H. Toward more proactive approaches to safety in the electronic health record era. Jt Comm J Qual Patient Saf 2017;43(10):540-547 [FREE Full text] [doi: 10.1016/j.jcjq.2017.06.005] [Medline: 28942779]

25. Terezakis SA, Harris KM, Ford E, Michalski J, DeWeese T, Santanam L, et al. An evaluation of departmental radiation oncology incident reports: anticipating a national reporting system. Int J Radiat Oncol Biol Phys 2013;85(4):919-923. [doi: 10.1016/j.ijrobp.2012.09.013] [Medline: 23265568]

26. Elkin PL, Beuscart-Zephir MC, Pelayo S, Patel V, Nøhr C. The usability-error ontology. Stud Health Technol Inform 2013;194:91-96. [Medline: 23941937]

27. Stenzhorn H, Schulz S, Boeker M, Smith B. Adapting clinical ontologies in real-world environments. J Univers Comput Sci 2008;14(22):3767-3780 [FREE Full text] [doi: 10.3217/jucs-014-22-3767] [Medline: 20390048]

28. El-Sappagh S, Franda F, Ali F, Kwak KS. SNOMED CT standard ontology based on the ontology for general medical science. BMC Med Inform Decis Mak 2018;18(1):76. [doi: 10.1186/s12911-018-0651-5] [Medline: 30170591]

29. Warm D, Edwards P. Classifying health information technology patient safety related incidents - an approach used in Wales. Appl Clin Inform 2012;3(2):248-257 [FREE Full text] [doi: 10.4338/ACI-2012-03-RA-0010] [Medline: 23646074]

30. Kaplan B, Duchon D. Combining qualitative and quantitative methods in information systems research: a case study. MIS Quarterly 1988;12(4):571-586. [doi: 10.2307/249133]

31. Paré G. Enhancing the rigor of qualitative research: application of a case methodology to build theories of IT implementation. Qual Rep 2002;7(4):1-34 [FREE Full text] [doi: 10.46743/2160-3715/2002.1966]

32. Garcia L, Quek F. Qualitative research in information systems: time to be subjective? In: Lee AS, Liebenau J, DeGross JI, editors. Information systems and qualitative research. Boston, MA: Springer; 1997:444-465.

33. Bowker GC, Star SL. Sorting Things Out: Classification and Its Consequences. Cambridge, MA: MIT Press; 1999.

34. Vuokko R, Vakkuri A, Palojoki S. Harmonization of Finnish vaccination data. Stud Health Technol Inform 2021;281:709-713. [doi: 10.3233/SHTI210264] [Medline: 34042668]

35. Palojoki S, Vakkuri A, Vuokko R. The European cross-border health data exchange: focus on clinically relevant data. Stud Health Technol Inform 2021;281:442-446. [doi: 10.3233/SHTI210197] [Medline: 34042782]

36. Mäkelä-Bengs P, Hämäläinen P, Kalliokuusi V, Vuokko R. Manual annotation of narrative patient charts – Finnish experiences related to a multilingual text corpus. FinJeHeW 2019;11(1-2):76-85 [FREE Full text] [doi: 10.23996/fjhw.70241]

37. Vuokko R, Mäkelä-Bengs P, Härkönen M. Improving health care service delivery with a national code service. Stud Health Technol Inform 2014;205:323-327. [Medline: 25160199]

38. Mäkelä-Bengs P, Vuokko R. Code service in social and health care: process description of the THL code service. National Institute for Health and Welfare. 2013. URL: https://core.ac.uk/download/pdf/18618461.pdf [accessed 2021-07-19]

39. Morris RD, Fletcher JM. Classification in neuropsychology: a theoretical framework and research paradigm. J Clin Exp Neuropsychol 1988;10(5):640-658. [doi: 10.1080/01688638808402801] [Medline: 3066798]

40. de Keizer NF, Abu-Hanna A, Zwetsloot-Schonk JH. Understanding terminological systems. I: terminology and typology. Methods Inf Med 2000;39(1):16-21. [Medline: 10786065]

41. Hancock D. The modernisation of statistical classifications in knowledge and information management systems. Electron J Knowl Manag 2017;15(2):126-144 [FREE Full text]

42. Borycki EM, Kushniruk AW, Bellwood P, Brender J. Technology-induced errors. The current use of frameworks and models from the biomedical and life sciences literatures. Methods Inf Med 2012;51(2):95-103. [doi: 10.3414/ME11-02-0009] [Medline: 22101488]

43. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform 2002;35(5-6):352-359 [FREE Full text] [doi: 10.1016/s1532-0464(03)00034-0] [Medline: 12968784]

44. Best practice guidelines for developing international statistical classifications. United Nations Statistics Division. 2013. URL: https://unstats.un.org/unsd/classifications/bestpractices/Best_practice_Nov_2013.pdf [accessed 2021-07-11]

Abbreviations

EHR: electronic health record
IT: information technology

XSL•FO
RenderX

Viewpoint

Cost and Effort Considerations for the Development of Intervention Studies Using Mobile Health Platforms: Pragmatic Case Study

Dan Thorpe[1], BMUS, MPhil; John Fouyaxis[1], BSc, BEng, Grad Dip; Jessica M Lipschitz[2], BA, PhD; Amy Nielson[1], BPsychSci (Hons); Wenhao Li[1], BSc, MSc, PhD; Susan A Murphy[3], BS, PhD; Niranjan Bidargaddi[1], BEng, PhD

[1]Flinders Digital Health Research Lab, College of Medicine and Public Health, Flinders University, Clovelly Park, Australia

[2]Brigham and Women's Hospital, Harvard, Boston, MA, United States

[3]Radcliffe Institute, Harvard University, Boston, MA, United States

Corresponding Author:
Dan Thorpe, BMUS, MPhil
Flinders Digital Health Research Lab
College of Medicine and Public Health
Flinders University
Tonsley Campus
1284 South Road
Clovelly Park, 5042
Australia
Phone: 61 0872215238
Email: dthorpe@flinders.edu.au

Abstract

Background: The research marketplace has seen a flood of open-source or commercial mobile health (mHealth) platforms that can collect and use user data in real time. However, there is a lack of practical literature on how these platforms are developed, integrated into study designs, and adopted, including important information around cost and effort considerations.

Objective: We intend to build critical literacy in the clinician-researcher readership into the cost, effort, and processes involved in developing and operationalizing an mHealth platform, focusing on Intui, an mHealth platform that we developed.

Methods: We describe the development of the Intui mHealth platform and general principles of its operationalization across sites.

Results: We provide a worked example in the form of a case study. Intui was operationalized in the design of a behavioral activation intervention in collaboration with a mental health service provider. We describe the design specifications of the study site, the developed software, and the cost and effort required to build the final product.

Conclusions: Study designs, researcher needs, and technical considerations can impact effort and costs associated with the use of mHealth platforms. Greater transparency from platform developers about the impact of these factors on practical considerations relevant to end users such as clinician-researchers is crucial to increasing critical literacy around mHealth, thereby aiding in the widespread use of these potentially beneficial technologies and building clinician confidence in these tools.

(JMIR Form Res 2022;6(3):e29988) doi:10.2196/29988

Introduction

Interest in integrating mobile technology into research has grown exponentially over the last 2 decades, particularly as a means of collecting real-time observational data and intervention outcomes [1]. A newer, but equally growing area of research involves Just-In-Time Adaptive Interventions (JITAIs) [2], which use data collected via mobile/wearable devices, such as symptoms and behaviors, to identify opportune moments for delivering personalized interventional content in real time.

These increasingly popular techniques have created a thriving marketplace for mobile health (mHealth) platforms designed to aggregate mobile app data, deliver intervention content, and observe outcomes. For example, the open-source AWARE

framework is an Android platform designed to address the lack of availability of open and reusable software for creating context-aware apps on mobile devices [3]. Similarly, Remote Assessment of Disease and Relapse-base is a scalable, fully functional, remote Internet of Things data collection platform for real-time remote sensor data collection [4]. Such platforms allow clinicians and researchers to collect (1) ecological momentary assessment (EMA) data, for example, active self-report surveys collected on a mobile or wearable device; (2) sensor data from external devices, such as data collected by a wrist-worn device like a Fitbit [5] or a mattress sensor like the Withings Sleep Analyzer [6]; and (3) sensor data directly collected from participants' smartphones, including step counts, geolocation data throughout the day, and so on.

Clinician-researchers seeking to integrate mobile app data into their research or practice either to deliver interventions or to monitor existing interventions have a plethora of options. However, very few apps include all these features, and implementing these apps on a large scale can be problematic without specific technical and implementation knowledge [4,7]. Should one wish to learn these skills, there is limited documentation with respect to understanding the pragmatic processes required for setting up these studies, building new features, and assessing the feasibility of technical solutions to these research questions. Furthermore, for many clinician-researchers, this is a decision that is outside of their realm of expertise. A wrong decision can dramatically impact the feasibility of meeting a given set of research aims. Indeed, despite the availability of several mHealth platforms, guidance for integrating intervention aims with technical capabilities—from a combined technical and implementation perspective—is difficult to find and would be valuable to researchers working with these tools.

To contribute to this body of knowledge, this paper describes the programmatic processes involved using our experiences with developing an mHealth platform, Intui. The mHealth platform is described with a particular focus on the decisions and choices of relevance to health services and clinician-researchers, including the procedure, cost and effort, and design choices at the product and project levels. To illustrate this further, we present the software results of these processes in action, namely a practical example of the adaptation of Intui to a behavioral activation therapy context, adding practical depth and detail to our procedures.

Methods

Materials

Intui Platform: Design Decisions and Features

The Intui platform is a configurable cloud-based service designed to support mHealth data collection and intervention studies (Figure 1). The 4 core software elements of the platform are (1) the Android and iOS smartphone apps for participant use; (2) lightweight cloud functions, also known as microservices, to handle reusable functional requirements provided in the app and the dashboard—for example, security, user management, communication, data management and data analysis—in isolation and at scale; (3) the web-based data dashboard for clinician-researchers to manage study participants and data; and (4) cloud storage on Amazon or Google data centers.

The app and dashboard user interfaces are implemented in a cross-platform app development framework, Ionic, which allows us to create and implement modular components that can be reused in multiple projects as required. Ionic also supports easy integration of third-party plug-ins, which can enable rapid development of new features [8]. For back-end functionality and data storage, Google and Amazon cloud services are used owing to their easy scalability and ability to integrate functionalities implemented in a number of programming languages as microservices.

Figure 1. Components, architecture, and example functionalities of the Intui platform.
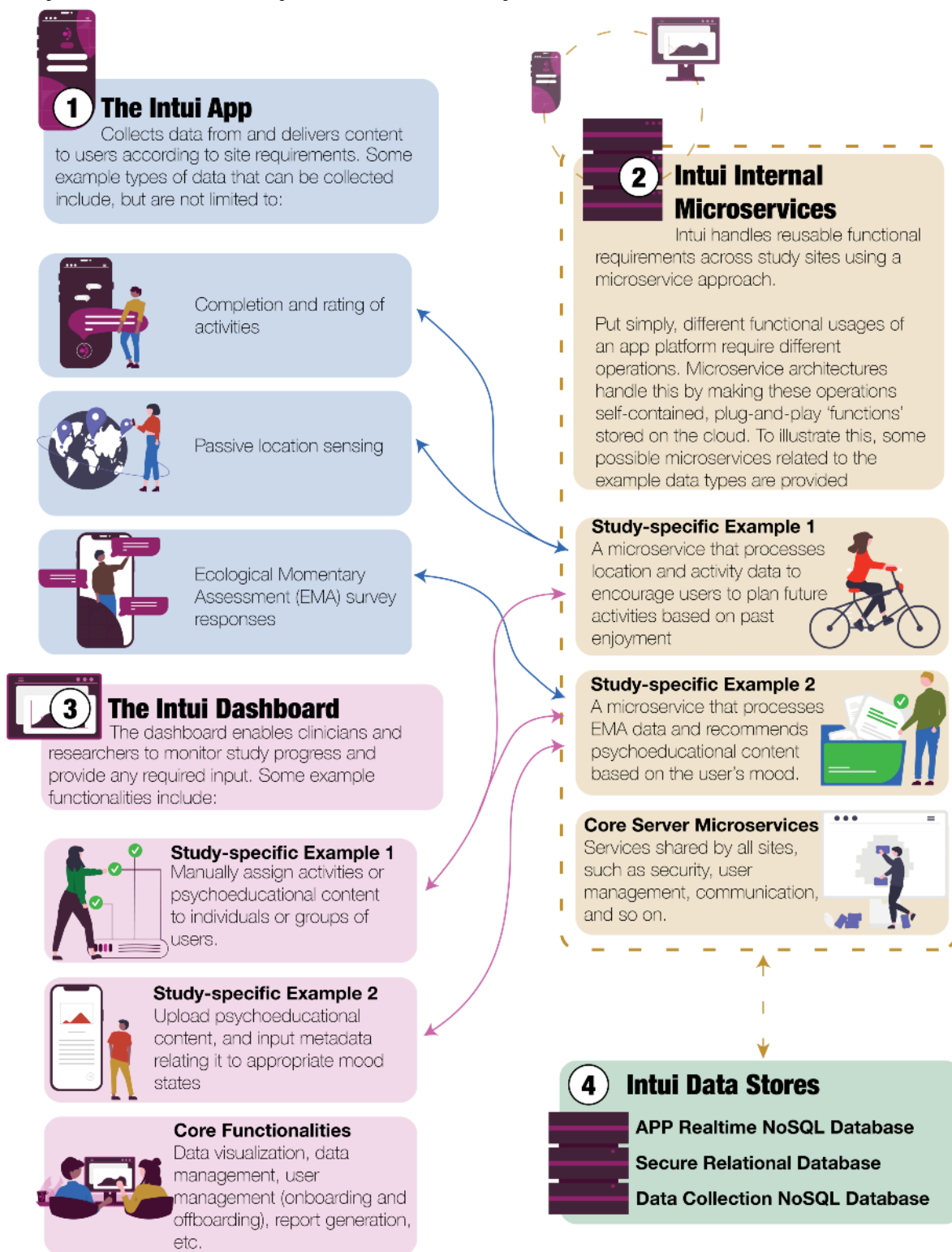
Study Design and Initiation

The process of initiating a new study within the Intui platform consists of three parts, guided by the Intui platform software development team: (1) setting up data hosting, (2) configuring Intui to the study-specific requirements, and (3) developing and implementing any custom study components.

Data Hosting Setup

The data hosting setup begins with an Intui software developer preparing the study-specific databases. The Intui platform is designed such that each study has a separate set of databases, 1 set for storing deidentified study data and another for securely storing personally identifiable participant details in an encrypted format. This provides an assurance that data segregation exists from 1 study to another for privacy and ethical considerations,

as discussed in detail later. At this stage, a unique study code is also generated for registered clinician-researchers.

Study-Specific Configuration

Once the hosting setup is complete, registered clinician-researchers receive a welcome email with details to access the web-based data dashboard, as well as the study code to share with prospective participants using the app. Researchers can also use the data dashboard to perform administrative functions, such as inviting participants, removing participants, downloading data, viewing summary statistics, and monitoring study progress.

Based on the unique requirements of each study, Intui software developers extract the study-specific configurations. These configurations define the inclusion of existing functionalities and components of the Intui platform as well as any parameters required for them to operate within the research design. For low-complexity studies, this includes the data collection requirements for phone sensor data and external device sensors, the definition of EMA question sets with branching logic, and protocol details such as the length of the study and timing of reminders related to the proposed data collection.

Development and Implementation of Any Custom Study Components

High-complexity studies often involve the implementation of custom study components. These components may include app user interface (UI) elements for intervention delivery, custom-built data gathering functionalities, and the implementation of microservices to handle the functional logic requirements of more complex study designs. Intui software developers work with researchers to define, design, and implement these components, or integrate those developed by third parties. This flexibility allows researchers to retain full software development control over aspects of more complex study designs.

In addition, UI components that encompass visual elements (styling and display), data (what is shown), and logic (what needs to be stored and any conditional functionality) may need to be developed. These can be designed by the Intui platform developers or by an external developer. Once a prototype UI component is available, the Intui platform developers integrate the component into the Intui app and release an updated version of the app with all added functionalities.

When a study has completed the initiation process, Intui software developers collaborate with the study researchers to run a short technical pilot. The pilot period validates the functional components and configured settings for the study ahead of participant recruitment. Technical pilots ensure that the right data are collected and that the study design has been correctly implemented to reduce the risk of participant dropout as a result of technical faults.

Measures

EMA Delivery

The design of the Intui mHealth platform permits the administration of various EMA question types and tailoring of the question flow and timing. For each study, the EMA question set(s) are defined within a study-specific configuration, and the schedule for question delivery is set up within the study protocol microservice of the Intui platform. EMA question types can include numerical sliders, Likert scales, single choice (eg, yes/no), multiple choice, and free-text input. Branching logic can be used to set conditions on which questions are shown. For example, if poor sleep is self-reported in the app, an additional question may appear asking for further information to differentiate difficulties in falling asleep.

Sensor Measures

The sensor data types and sources collectible through Intui are, essentially, unlimited. Sensors can be integrated with Intui at the application programming interface (API) level and through manual data entry methodologies, for example, participants entering results from an analog medical device.

Procedure

Participant Onboarding

Study participants are invited to download the Intui app from the respective Apple-iOS or Google-Android app stores. Participants register on the Intui app using the provided study code. The unique study code enables the study-specific configuration and features within the app and may also be used to link or segregate participant data. For example, a study may require the use of a patient app that feeds data into a clinician dashboard. In this scenario, clinicians are assigned a unique study code that they provide to participants, thereby linking multiple patient apps to an individual clinician's dashboard. This ensures that only authorized clinicians can access patient information linked to their study code, maintaining data segregation.

Collection of EMA Data

The timing of administering EMAs may be based on time-contingent, event-contingent, or answer-contingent schedules. Time-contingent schedules are configured at a predetermined time of the day/week (for example, a survey getting sent at 2 PM every Monday) or within a range of possible times (for example, a prompt being set to be sent at some point between 9 AM and 3 PM). Event-contingent schedules deliver a tailored prompt in response to characteristics, trends, or threshold values identified in the collected data (when event X occurs, nudge Y). Event-contingent nudges can be delivered in response to active data (when a self-reported threshold value is recorded or a trend is observed over several days) or passive data (like arriving at a specific location, reported through a smartphone's GPS sensor). The type of questions posed in the app and the schedule for delivering those questions can be customized to the requirements of each specific study. Additionally, the appearance of questions may be conditionally based on responses to previous questions (eg, if response is A, ask B). For example, if poor sleep is self-reported in the app, an additional question may appear asking for further information to differentiate difficulties in falling asleep.

The EMA question set(s) are defined within study-specific configuration settings and the schedule for question delivery is

set up within the study protocol microservice of the Intui platform for each study.

Delivery of JITAIs

Access to data in real time is critical to the implementation of JITAIs, as is the development of the necessary UI components to enable this implementation. External and internal programmers can access data gathered in real time through the secure data RESTful APIs in Intui. This can facilitate rapid and collaborative intervention development, as programmers can prototype approaches to process data in any environment (Python, R-Studio, Java, JavaScript) and schedule algorithms to run as independent applications (microservices) that use the same Intui database to read data and store processed results. Programmers can also use the Intui API, which can control push notifications to users and set up reminders to access interventions within the app. Such interventions can include brief text within push notifications, multimedia content within the app, or a set of EMA questions.

Data Extraction for Ongoing Analysis

The collected participant data may be downloaded at any time from the clinician-researcher dashboard or using the Intui API. Clinician-researchers can download data in the CSV format for offline analysis at any time from the data dashboard. Alternatively, separate data analytic applications can be implemented to programmatically read and process the gathered study data through APIs in real time and initiate study features (eg, present new content or messages to participants).

Considerations Unrelated to the Study Design

Support

Researchers can expect to receive stable support for the duration of their study from the Intui platform development team. Owing to the monthly support and maintenance fee, researchers (and study participants) can submit bug reports and expect support and resolution within reasonable time frames to ensure the continuous and uninterrupted progress of their study. The Intui platform development team includes app and cloud computing software developers who hold grant-funded project positions and are also working in the industry. These developers work with the lab on 1 or more projects in an ongoing manner.

Privacy and Ethics

Researchers need to establish an agreement with Flinders University to use the Intui platform that will outline the terms of hosting and sharing of data between the Intui development team and researchers. All data collected by the Intui app are transferred to and stored in 2 databases to lower the impact of a potential data security breach "at rest." Personal identifiable data (eg, names, email addresses used for registration, and logins) are stored in an encrypted user management database in Amazon Web Services data centers. Data collected and used in the study such as questionnaire/survey data and passive data streams are stored on Google Cloud Firestore. Due to the possibility of identifying individuals through their location data coordinates, all raw location data are encrypted before leaving the phone and remain encrypted during storage within the database in a format not readable by humans.

Cost and Effort

In the interest of transparency, we have provided the following example data in Australian Dollars (AUDs) associated with previous projects. The following conversion rate was used: AUD $1=US $0.72. In short, setup costs depend on the functionalities required by the study, the associated software development effort required for the app, and for any back-end development required on the functional database. This is further influenced by the complexity of the project, its duration, and the number of participants.

A low-complexity study that makes use of existing functionalities and components incurs the standard Intui platform study setup effort of 12 hours, roughly costed at AUD $100 per hour for the example in this paper, leading to a total of AUD $1200. This includes advice on study design and consultation with the team to determine study requirements. The baseline costs associated with technical support and maintenance efforts are 3 hours per month (approximately AUD $300) for studies involving up to 2000 participants.

In addition to these labor hours, infrastructure costs—like data hosting and cloud services—start from an AUD $100 per month allocation for studies of up to 2000 participants. Infrastructure costs for serverless environments are often difficult to estimate a priori but are based here on the Google Cloud Platform hosting costs at the time of printing (2021). Free credit tiers may influence infrastructure costs, for example, Cloud Function free tier of 2 million invocations per month, and Firebase free tier of 50,000 reads and 20,000 writes per day. The baseline allocation set here is used to cover Google Cloud resources used by App Engine for storage, logging, and reading operations, or where Cloud Function and Firebase usage exceeds free tier levels on heavy use days; however, this should be monitored closely during the early stages of any project to adjust for redundant usage and identify infrastructure cost savings at scale.

For a practical example involving a simple study, Intui was used to collect third-party device data over a 6-month observational study investigating the relationship between mental health and device data of at-risk young adults for a total cost of AUD $3000 [7]. In this study, participants installed the Intui app on their phone that extracted daily, passively recorded active time, and administered EMAs addressing participants' mood, sleep, and eating habits.

Higher complexity studies are costed on a time and effort basis. In this context, complexity can include the development and integration of custom UI elements, developing and ensuring high-quality integration of microservices that can handle custom intervention logics, or any other request that requires substantial effort beyond the hours outlined in the more basic worked example. The case study presented later in this paper provides a good worked example of the costs associated with a more complex project. Additional costs may also be incurred for alterations to the study design and data collection while a study is running, technically referred to as change requests. Studies with over 2000 participants incur additional monthly costs to account for more extensive support, maintenance, and infrastructure requirements.

Results

App Implementation Using Intui

To make these considerations less abstract, we illustrate how the Intui platform was used to implement an app to support a behavioral activation (BA) intervention. BA is an evidence-based technique used predominantly to manage depressive symptoms. The aim of BA is to increase engagement in behaviors that promote feelings of pleasure and mastery [9]. This technique requires participants to keep an accurate record of activities undertaken over a period and report their mood while undertaking that activity. The goal of this technique is for participants to learn to identify and schedule enjoyable and mastery-oriented activities. However, compliance is a recognized barrier to this treatment method, and few participants complete the prescribed protocol [10].

We implemented a BA app within the Intui platform with adaptive interventions to reduce the burden of monitoring and planning (Figure 2). This was achieved through adaptation of existing components and through the development of two new interventional components: (1) a nudging intervention to improve recall, planning, and adherence for self-monitoring; and (2) machine learning–assisted activity planning.

Figure 2. Intui behavioral activation implementation, including an activity recommendation (left) and a planning/logging (right) interface.
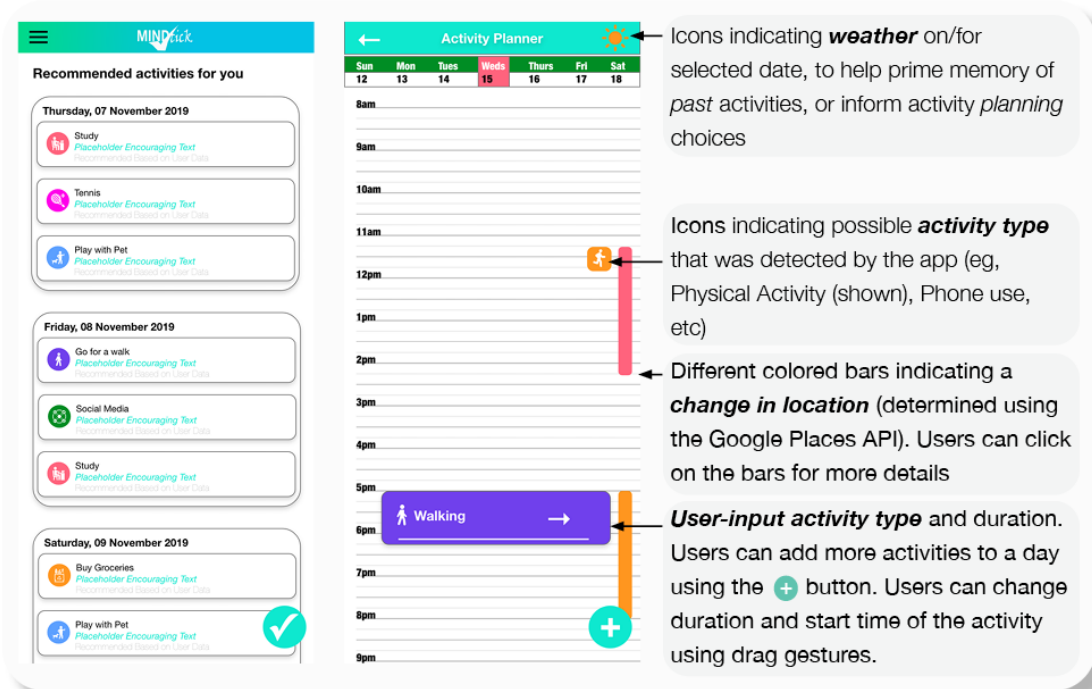
Study Design and Initiation

Data Hosting Setup

We configured a study-specific database on Google Firestore. User-reported data are stored in and recommended activities are fetched from the same study database. Additionally, we configured the app to collect location coordinates every few minutes and store them in the study database. We used a secure data API to read from and write to the study database. Next, we implemented 2 internet-based algorithms (described below) to read from and write to the study database to operationalize the 2 real-time intervention components.

Study-Specific Configuration

Pre-existing EMA components within Intui, as well as much of the back end and UI components of the app and dashboard, were reused or adapted for this project.

Development and Implementation of Any Custom Study Components

The BA app offered 2 interventional components using real-time data, namely nudging to improve recall and adherence to self-monitoring, and machine learning–assisted activity planning. These interventional components were designed specifically for this study, and they were crucial to its success. Outside of the algorithm development, the implementation and development of the custom app UI calendar elements, app UI location context detail elements, and the implementation of microservices to handle the logic for the recommender algorithm as well as the location clustering algorithm also contributed substantially to the development effort.

Interventions Developed

The first intervention component aims to enhance the user's ability to retrieve previously experienced activities and moods as part of self-monitoring. It achieves this by nudging relevant spatial and behavioral cues, involving push notifications and

in-app cues to predictably alter user behavior in a way that is predictable but without specifically forbidding any options or producing disincentivizing effects [11]. Using a calendar or list view interface in the app, users keep a detailed log of their past activities and associated moods. When a user clicks in the calendar to make an activity entry for any given day and time, contextual cues mined from the mobile phone location and usage data are displayed. The cues, shown as visual elements (Figure 2), correspond to where the users were (derived from location data, as described below), what activities they were doing (derived from movement sensor data), and the environment (weather of the location); all these are possible, thus aiding more accurate recall [12,13].

For cueing users to recall their activities and mood more accurately when they were recording data, we displayed visited places as vertical bars in the daily calendar view. Each bar visualized the start and end times of 1 visited place (Figures 2-3). When the bar was clicked, the users were able to see a map and the details of each location, including the precise start and end times. These vertical bars were generated daily, using the results of a k-means clustering algorithm that was set to run at midnight. This algorithm works by taking a continuous stream of location coordinates gathered from the phone sensors, clustering them into locations, and then outputting them as a list of visited places. Entries for the start and end times of each location, and the central GPS point of the cluster marked as the visited place were made into the database. Once users confirm the type of activity they have completed, they are asked to evaluate the activity through a short questionnaire (Figure 3).

The second intervention component aims to assist with planning activities by making personalized activity recommendations learned from the users' self-monitored data and that of other users with similar behavioral patterns. The planner interface displays a ranked order list of activity recommendations drawn from a mix of previously recorded activities and random activities (Figure 2). Users select an activity and schedule when to undertake it in the future. Users' responses to the suggested activities and their evaluation are processed through a bilinear model to refine the ranked order listing of future recommendations [14]. Over time, the recommender interface personalizes the ranked order list with higher-order ranking for activities that reflect long-term user preferences; that is, activities that have resulted in better mood outcomes are the most likely ones to be completed.

To ensure that activity suggestions were personalized to user circumstances, the order of activities displayed in the planner view were dynamically ranked based on past data (Figure 3). To achieve this, we implemented the linear upper confidence bound algorithm, which weighs activities reaping the most rewards (in terms of mood and likelihood to be completed) when exploring new options that may be equally or more rewarding; this is a very important consideration in the context of BA, where targeting reward responsiveness is crucial [15]. The algorithm takes user activity preferences from past data entries and a batch model trained using the first 8 weeks of data from all the users, and it outputs a ranked list of activities to help users prioritize activities that may reward them. A second microservice was scheduled to refresh the batch model daily at midnight after an 8-week period.

Figure 3. Intui behavioral activation app activity evaluation questionnaire. After inputting an activity or clicking on a machine learning–generated activity suggestion in the planner view (Step 1), users are asked to confirm the details of the activity (Step 2), and then answer 3 short questions evaluating the activity (Step 3). API: application programming interface.
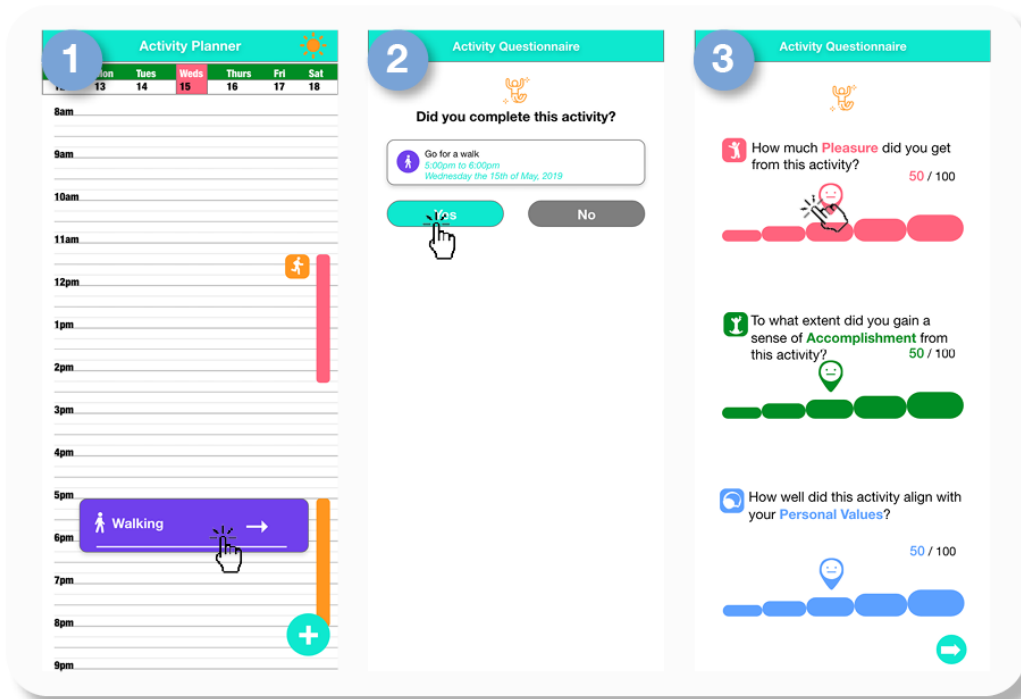
Cost and Effort

These algorithms were implemented in Python and run as independent microservices on an Ubuntu instance. This work was carried out independently by a research assistant with a background in data science and Python programming. The specifications for storing the results of the algorithms were established by the platform developers. Additionally, these platform developers designed and implemented the UI elements that incorporated the outputs of these algorithms into the clinician-researcher dashboard. This case study was completed in 62 hours (costed at AUD $100 per hour for $6200).

Discussion

Principal Findings

The clinician-researchers seeking to integrate mobile app data into their research either to deliver interventions or to monitor existing interventions have a plethora of options. As evidenced by the overall process of developing Intui and the case study outlined in this study, the considerations clinician-researchers must bear in mind when setting out to perform this research are extensive. However, Intui represents just 1 approach to designing, developing, and implementing these tools, and regardless of the platform, there is a need to reflect on the ability of clinician-researchers and their teams to meet the logistical and technical requirements underpinning the success of these platforms in their contexts.

Ecological studies, especially longitudinal studies, often require extensive infrastructure and information technology support that demand careful planning to ensure high-quality development, implementation, and maintenance [16]. The frequency at which extremely granular data are stored, received, segmented, analyzed, and presented is also a challenge that can be underestimated, thus affecting the ability of platforms to be used for multiple research outcomes. All these decisions drastically impact the feasibility of delivering projects, are difficult to plan for and cost without a nontechnical background, and can either limit or enable the possibilities of a project. Indeed, this is critical to the provision of the support we have integrated into our operations as nonnegotiable for Intui.

These issues are not confined to our experiences in developing Intui and supporting trials using it; indeed, cost and effort—including the up-front and opportunity costs due to the perceived loss of time in learning new tools—training, and support, were listed as prescient concerns by de Grood and colleagues in their scoping review of eHealth technology adoption by physicians [17]. Open-sourcing projects claims to address these concerns in many respects, namely through literal transparency by sharing the bare bones of software for all to scrutinize [18]. In an age of WebMD self-diagnoses, we are assured that our medical audience requires little convincing that open-sourcing data only results in transparency of the pragmatic type if it is received by an audience able to understand, contextualize, and critique it [7,19,20]. In the same sense that health consumers are ultimately justified in their desire to combat medical gatekeeping and the harm it can cause [7,19], health professionals have every right to question the authority and hegemony of software developers and technology companies on eHealth, especially given the questionable quality of these software packages sometimes [21-23]. However, the solution in both cases lies on a middle ground; solutions can be developed through new models with transparent collaboration and through the development of a critical audience capable of engaging with these digital realities in a reflexive and an ongoing manner [7,20,24]. Through providing insights into the cost, effort, and processes involved in developing and operationalizing Intui in this paper, we hope to have contributed to the development of this critical audience as well as to more transparent norms in mHealth platform development and operationalization in a broader sense.

Conclusions

Enabling researchers to rapidly implement mobile data collection and intervention studies can ultimately lead to increased understanding and access to evidence-based behavioral and mental health interventions. Based on a case study involving our own platform and on critical reflection, this paper has outlined some practical considerations for those hoping to engage in these technologies. Greater transparency in terms of the cost, effort, and support across the board in this field will only enable the widespread use of these beneficial methodologies and build clinician confidence in these tools.

Conflicts of Interest

None declared.

References

1. Perry B, Herrington W, Goldsack JC, Grandinetti CA, Vasisht KP, Landray MJ, et al. Use of mobile devices to measure outcomes in clinical research, 2010-2016: a systematic literature review. Digit Biomark 2018 Jan;2:11-30 [FREE Full text] [doi: 10.1159/000486347] [Medline: 29938250]
2. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr;16:31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
3. Ferreira D, Kostakos V, Dey AK. AWARE: mobile context instrumentation framework. Front ICT 2015 Apr;2:1-9. [doi: 10.3389/fict.2015.00006]

4. Ranjan Y, Rashid Z, Stewart C, Conde P, Begale M, Verbeeck D, The Hyve, The RADAR-CNS Consortium. RADAR-Base: open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, and mobile devices. JMIR Mhealth Uhealth 2019 Aug;7(8):e11734 [FREE Full text] [doi: 10.2196/11734] [Medline: 31373275]

5. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. JMIR Mhealth Uhealth 2018 Aug;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]

6. Edouard P, Campo D, Bartet P, Yang RY, Bruyneel M, Roisman G, et al. Validation of the Withings Sleep Analyzer, an under-the-mattress device for the detection of moderate-severe sleep apnea syndrome. J Clin Sleep Med 2021 Jun;17(6):1217-1227. [doi: 10.5664/jcsm.9168]

7. Stern G. Self-diagnosing: on the proper role of sites like WebMD. The Atlantic. URL: https://www.theatlantic.com/health/archive/2012/10/self-diagnosing-on-the-proper-role-of-sites-like-webmd/263297/ [accessed 2021-12-02]

8. Rieger C, Majchrzak TA. Towards the definitive evaluation framework for cross-platform app development approaches. J Syst Softw 2019 Jul;153:175-199. [doi: 10.1016/j.jss.2019.04.001]

9. Cuijpers P, van Straten A, Warmerdam L. Behavioral activation treatments of depression: a meta-analysis. Clin Psychol Rev 2007 Apr;27(3):318-326. [doi: 10.1016/J.CPR.2006.11.001]

10. Kazantzis N, Brownfield NR, Mosely L, Usatoff AS, Flighty AJ. Homework in cognitive behavioral therapy: a systematic review of adherence assessment in anxiety and depression (2011–2016). Psychiatr Clin North Am 2017 Dec;40(4):625-639. [doi: 10.1016/j.psc.2017.08.001]

11. Anderson J. Nudge: Improving decisions about health, wealth, and happiness, Richard H. Thaler and Cass R. Sunstein. Yale University Press, 2008. x + 293 pages. [Paperback edition, Penguin, 2009, 320 pages.]. Economics and Philosophy 2010 Oct;26(3):369-376. [doi: 10.1017/S0266267110000301]

12. Rabbi M, Li K, Yan HY, Hall K, Klasnja P, Murphy S. ReVibe: a context-assisted evening recall approach to improve self-report adherence. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Dec;3(4):1-27 [FREE Full text] [doi: 10.1145/3369806] [Medline: 34164595]

13. Robin J, Moscovitch M. Familiar real-world spatial cues provide memory benefits in older and younger adults. Psychol Aging 2017 May;32(3):210-219. [doi: 10.1037/pag0000162] [Medline: 28230384]

14. Zhang T, Jarrad G, Murphy SA, Bidargaddi N. A smartphone-based behavioural activation application using recommender system. In: UbiComp/ISWC '19 Adjunct: Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers. 2019 Sep Presented at: UbiComp '19: The 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 9-13, 2019; London, United Kingdom p. 250-253. [doi: 10.1145/3341162.3343785]

15. Forbes CN. New directions in behavioral activation: using findings from basic science and translational neuroscience to inform the exploration of potential mechanisms of change. Clin Psychol Rev 2020 Jul;79:101860. [doi: 10.1016/j.cpr.2020.101860] [Medline: 32413734]

16. Burke LE, Shiffman S, Music E, Styn MA, Kriska A, Smailagic A, et al. Ecological momentary assessment in behavioral research: addressing technological and human participant challenges. J Med Internet Res 2017 Mar;19(3):e77 [FREE Full text] [doi: 10.2196/jmir.7138] [Medline: 28298264]

17. de Grood C, Raissi A, Kwon Y, Santana MJ. Adoption of e-health technology by physicians: a scoping review. J Multidiscip Healthc 2016 Aug;9:335-344 [FREE Full text] [doi: 10.2147/JMDH.S103881] [Medline: 27536128]

18. Torous J, Wisniewski H, Bird B, Carpenter E, David G, Elejalde E, et al. Creating a digital health smartphone app and digital phenotyping platform for mental health and diverse healthcare needs: an interdisciplinary and collaborative approach. J Technol Behav Sci 2019 Apr;4(2):73-85 [FREE Full text] [doi: 10.1007/s41347-019-00095-w]

19. Shen C, Nguyen M, Gregor A, Isaza G, Beattie A. Accuracy of a popular online symptom checker for ophthalmic diagnoses. JAMA Ophthalmol 2019 Jun;137(6):690-692 [FREE Full text] [doi: 10.1001/jamaophthalmol.2019.0571] [Medline: 30973602]

20. Kemper J, Kolkman D. Transparent to whom? No algorithmic accountability without a critical audience. Inf Commun Soc 2018 Jun;22(14):2081-2096. [doi: 10.1080/1369118X.2018.1477967]

21. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020 Jan;3:4 [FREE Full text] [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]

22. Nouri R, Niakan Kalhori SR, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018 Aug;25(8):1089-1098 [FREE Full text] [doi: 10.1093/jamia/ocy050] [Medline: 29788283]

23. Buijink AWG, Visser BJ, Marshall L. Medical apps for smartphones: lack of evidence undermines quality and safety. Evid Based Med 2013 Jun;18(3):90-92. [doi: 10.1136/eb-2012-100885] [Medline: 22923708]

24. Karnieli-Miller O, Eisikovits Z. Physician as partner or salesman? Shared decision-making in real-time encounters. Soc Sci Med 2009 Jul;69(1):1-8. [doi: 10.1016/j.socscimed.2009.04.030] [Medline: 19464097]

Abbreviations

API: application programming interface
BA: behavioral activation
EMA: ecological momentary assessment
JITAIs: Just-In-Time Adaptive Interventions
mHealth: mobile health
UI: user interface

XSL•FO
RenderX

Original Paper

Resilience in 2021—Descriptive Analysis of Individuals Accessing Virtual Mental Health Services: Retrospective Observational Study

Grant Graziani[1], PhD; Sarah Kunkle[1], MPH; Emily Shih[1], PhD

Ginger, San Francisco, CA, United States

Corresponding Author:
Sarah Kunkle, MPH
Ginger
116 New Montgomery Street
San Francisco, CA, 94105
United States
Phone: 1 7175197355
Email: sarah@ginger.io

Abstract

Background: Psychological resilience has been extensively studied by developmental researchers, and there is a growing body of literature regarding its role in psychiatry and psychopathology research and practice. This study contributes to this growing literature by providing real-world evidence on the relationship between resilience and clinical symptoms among a large sample of employed Americans.

Objective: This study aimed to describe resilience levels in individuals accessing Ginger, a virtual mental health system, in addition to the association of resilience with demographic characteristics, baseline depression, and anxiety symptoms.

Methods: We conducted a retrospective observational study of 9165 members who signed up for Ginger and completed a baseline survey between January 1 and August 5, 2021. We used multivariate regression models to test for associations between baseline resilience and other member characteristics.

Results: Baseline resilience scores centered on a mean of 23.84 (SD 6.56) and median of 24 (IQR 8) out of 40, with 81.0% (7424/9165) of the sample having low resilience at baseline. Despite having relatively higher resilience scores, members with no or mild depression or anxiety still had low resilience scores on average. Self-reported suicidal ideation was associated with lower resilience.

Conclusions: Overall, members had low baseline resilience, similar to resilience levels observed in trauma survivors in prior studies. Younger members and those with higher levels of depression and anxiety at intake reported lower levels of resilience at baseline. Notably, members with no or mild depression or anxiety still had low resilience scores on average, suggesting a need for mental health support among individuals who might not typically be recommended for treatment based on traditional clinical assessments, such as the 9-item Patient Health Questionnaire (PHQ-9) and the 7-item Generalized Anxiety Disorder scale (GAD-7). Two suggestions for topics of future research are to develop treatment recommendations based on the Connor-Davidson Resilience Scale and to understand the interaction between resilience levels and symptom-based outcome measures, such as the PHQ-9 and the GAD-7.

Introduction

Resilience and Adaptability

Psychological resilience, subsequently referred to as resilience, represents the personal qualities that enable an individual to thrive in the face of adversity. It can be viewed as a measure of the ability to cope with stress and is potentially an important target of treatment for anxiety, depression, and stress reactions, such as posttraumatic stress disorder (PTSD). Research has shown that resilience is a multidimensional characteristic and that it may vary with context, time, age, gender, and cultural origin, as well as different life circumstances [1,2].

The construct of resilience has long been of interest to developmental psychologists [3]. Considered a personal strength, resilience can contribute to positive functioning and optimal development and can prevent negative emotions, thoughts, and behaviors [4]. While it does not imply total invulnerability to the development of psychiatric disorders, resilience serves as a protective factor against the development and onset of psychopathology [3]. An evaluation of hypotheses about the relationship of resilience to personality traits, coping, and psychiatric symptoms found that it demonstrates strong positive correlations with extraversion and conscientiousness, and it can moderate the relationship between retrospective reports of childhood emotional neglect and current psychiatric symptoms [3]. Researchers have also pointed out that resilience is not necessarily a feature of an individual's internal psychological processes but can be a product of an individual's social and psychological ecosystem, including individual, family, community, and cultural factors [5].

There is a growing body of literature on resilience in the fields of psychiatry and psychopathology, which tends to be more focused on disease and pathology. A new approach in psychopathology research, advocated by some authors, focuses on positive adaptation in response to stress [3]. As part of this approach, there is interest in moving beyond an emphasis on pathology and focusing on prevention through human strengths and protective factors [6]. The importance of resilience as an inherent ability to manage daily stresses, as well as to overcome severe trauma, has increasingly been recognized [7]. Studies of frontline health care workers during the beginning of the COVID-19 pandemic and responders to the September 11, 2001, attacks have shown that resilience is important for both risk of PTSD and the ability to overcome trauma [8,9]. Further, researchers have made progress modeling the neurobiological components of resilience, pointing to both physiological processes and genetic factors that shape a person's resilience [10].

COVID-19

The COVID-19 pandemic presented the need for a prolonged period of social distancing and potential isolation, a changing and uncertain time frame for improvement in conditions and lifting of restrictions, and uncertain political and economic implications. Recent studies have shown that individuals with lower resilience scores experienced increased odds of mental distress and expressed greater difficulty coping with the emotional challenges of the pandemic crisis [11,12]. Given these unique challenges to resilience and mental well-being, more attention is being given to increasing resilience from both research and clinical perspectives. In particular, health systems have recognized the need to promote resilience among both health care workers and patients. As New York City became the epicenter of the pandemic in the United States, the Mount Sinai Health System created the Center for Stress, Resilience, and Personal Growth in order to address the pandemic's psychological impact on the health care workers in the system; they also created a resilience app, the Wellness Hub, as a standalone digital platform offering users a suite of tools that they could interact with on a daily basis [13]. Furthermore, the pandemic has spurred increased interest in not only

understanding the neurobiological and cultural process that shape resilience but also in developing interventions that improve resilience, as individuals around the globe cope with the implications of the COVID-19 pandemic [14]. In this regard, digital health platforms can play an important role in delivering such interventions.

Study Objectives

Given current events, increased demand for mental health services, and newer disciplines like behavioral health coaching, it is important to understand individual needs, particularly those that might not be captured in traditional clinical assessments, such as the 9-item Patient Health Questionnaire (PHQ-9) and the 7-item Generalized Anxiety Disorder scale (GAD-7). With this in mind, Ginger, an on-demand mental health system, began collecting self-reported resilience data from its members beginning in December 2020. These data are used in this study, which aims to describe resilience levels in individuals accessing virtual mental health services, a population that is less understood given the relatively nascent industry. In particular, we aim to answer the following research questions: (1) What is the distribution of baseline resilience? and (2) To what extent are baseline scores correlated with demographic characteristics and concurrent depression and anxiety symptoms?

Methods

Overview

This is a retrospective observational study of individuals who accessed Ginger. Data were collected from Ginger members between January 1 and August 5, 2021.

The Ginger System

Ginger provides virtual on-demand mental health services, primarily through employee or health plan benefits. Via a mobile app platform, Ginger members can access behavioral health coaching, teletherapy, and telepsychiatry, as well as self-guided content and assessments. This system has been described in more detail in prior publications evaluating depression and anxiety outcomes as measured by the PHQ-9 and the GAD-7 [15,16].

Participants

Study participants had access to the Ginger system as part of their employer or health plan benefits. Internal clinical protocols include the following exclusionary criteria where self-directed telehealth is likely not appropriate and where more specialized and urgent psychiatric services are required:

1. Active suicidal ideation.
2. Active high-risk self-harm behavior.
3. Two or more hospitalizations within the past 6 months or one hospitalization in the past month for psychiatric reasons.
4. Certain symptoms of psychosis that are poorly managed (eg, member is not medication-compliant or symptoms are unresponsive to treatment) and are likely incompatible with telehealth.
5. A primary diagnosis of a substance use disorder or moderate to severe substance abuse issues, due to the high complexity,

severity, and risk frequently associated with such members, as well as the need for specialized care.

6. Active eating disorders with symptoms considered to be high risk.

7. Ongoing grave disability, including certain patients who are bipolar with active mania, hypomania, or mixed episodes; are unmedicated; or have poor compliance with a medication regimen over time.

8. Two or more medical hospitalizations in the last month, due to the high likelihood that the individual has a poorly controlled medical condition that requires close monitoring.

For this study, we included Ginger users aged 18 years or older who joined during the study data collection period.

Data Collection

As part of its measurement-based care system, Ginger uses various assessments, including the PHQ-9 and the GAD-7. Since December 2020, Ginger has used the 10-item Connor-Davidson Resilience Scale (CD-RISC 10), also referred to as an adaptability check-in, to track progress beyond depression and anxiety symptoms. This is particularly relevant to understand needs of "subclinical" members (ie, those members who screen negative for depression or anxiety at intake). This measure was selected due to behavioral health coaching's focus on building resilience and its strength-based focus in contrast to more traditional symptom measures, such as the PHQ-9 and the GAD-7. The CD-RISC 10 was sent to members 1 week after they signed up, and a follow-up survey was sent to members every 30 days. Importantly, members who signed up but did not engage with the app past the 1-week mark did not complete the baseline survey. In this way, members with a low likelihood of meaningful engagement, which is a proxy for behavioral health need, were excluded from the sample. Visuals of how this survey appeared to members are shown in Figure 1.

Figure 1. Screenshots of the Ginger mobile app showing the 10-item Connor-Davidson Resilience Scale.
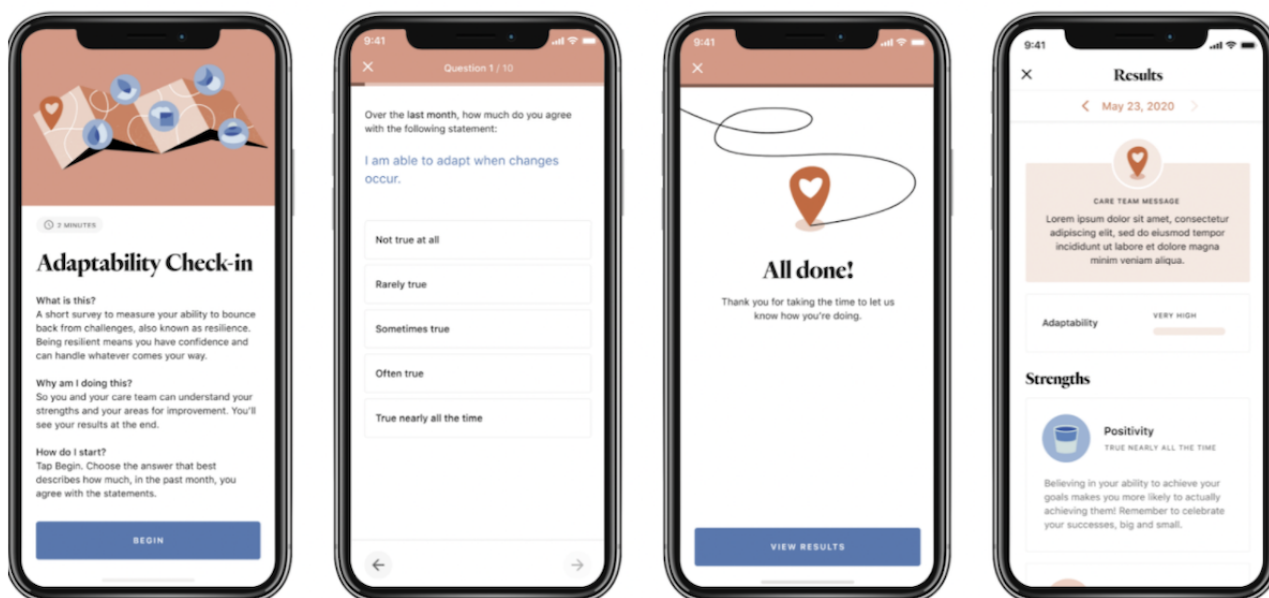
Measures

The Connor-Davidson Resilience Scale

As mentioned above, Ginger uses the CD-RISC 10 to measure self-reported perceived resilience. The development of the Connor-Davidson Resilience Scale (CD-RISC) [1] arose from the researchers' extensive treatment of individuals suffering from PTSD. They initially developed a 25-item scale to measure resilience, or how well one is able to adapt to change and bounce back after stressful events, tragedy, or trauma. Two briefer versions, the CD-RISC 10 [17] and the 2-item CD-RISC [18], were subsequently developed by other research teams. The CD-RISC 10 has demonstrated robust validity, reliability, and practicality [1]. Since its development in 2003, the CD-RISC has been translated into many different languages and studied in a variety of populations [1].

The CD-RISC 10 contains 10 of the original 25 items from the CD-RISC. The 10 topics included in the CD-RISC 10 are as follows: confidence, determination, flexibility, focus, grit,

perseverance, personal growth, positivity, self-reliance, and weathering emotions. For each of the 10 items, respondents were asked to select one of the following responses to a statement (eg, "I am able to adapt when changes occur"): not true at all (0), rarely true (1), sometimes true (2), often true (3), and true nearly all the time (4). A respondent's total score could range from 0 to 40. Results from the US population indicated that the quartiles for this measure are as follows: quartile 1, 0 to 29 points; quartile 2, 30 to 32 points; quartile 3, 33 to 36 points; and quartile 4, 37 to 40 points [17].

Baseline Characteristics

For each member, the following data were either collected at baseline or were fixed characteristics of members: age group, gender, geographic region, PHQ-9 score, and GAD-7 score. The demographic and location data were not self-reported. Instead, they were reported by a member's parent organization, which was either their employer or their health insurance plan. The baseline PHQ-9 and GAD-7 data were collected within the Ginger system. Baseline PHQ-9 and GAD-7 scores were

selected by looking at the window from 1 week before to 1 week after a member's baseline CD-RISC 10 score was collected and choosing the first PHQ-9 and GAD-7 scores in that window.

For many Ginger members in this study, baseline characteristics were missing. Data were missing due to one of two reasons. First, a member's parent organization may not have shared the member's demographic information. Thus, missing demographic data is a signal of a member's parent organization and not necessarily a signal of information specific to a given member. For example, of the 249 parent organizations represented in this study, 116 (46.6%) reported all of their members' gender information and 121 (48.6%) did not report their members' gender information. The remaining 12 (4.8%) reported gender information for some but not all of their members. Second, there may not have been a PHQ-9 or GAD-7 score within the 1-week window around the collection of a member's baseline CD-RISC 10 score. This could be due to a member not completing the PHQ-9 or the GAD-7 at all or due to the timing of their completion of surveys falling outside the 2-week window.

Analyses

Sample

This study included 9165 Ginger members who completed a baseline survey at any point from January 1 to August 5, 2021. This sample will be referred to as the baseline sample.

Summary Statistics and Subgroup Analysis

Our descriptive analysis summarized baseline resilience scores and presented the mean (SD) and median (IQR) of baseline scores. We used a Welch t test to analyze differences in means across subgroups of members with unequal variances when a category had two groups (eg, gender). For categories with more than two groups (eg, census regions), we used an F test as part of an analysis of variance to test for significant differences in means across the groups. Further, to understand whether members with missing data had significantly different outcomes than those without missing data, we performed Welch 2-tailed t tests comparing means across the two groups.

Descriptive Multivariate Regressions

In order to understand the associations between resilience and specific covariates, we leveraged a multivariate regression model as part of our descriptive analysis of baseline scores. This methodology accounted for possible correlations among covariates and isolated the relationship between each covariate and resilience, holding all other covariates constant. We estimated an ordinary least squares (OLS) linear regression predicting baseline scores. The following categorical independent variables were included in the model: gender, age group, census region, and indicators for whether a member's baseline PHQ-9 and GAD-7 scores were each above 10. For each of these independent variables, a category for members with missing data was included. No interaction terms were included, although, theoretically, there could be significant differences at a more granular level. Additionally, self-reported suicidal ideation (ie, indicating more frequently than "never" on question 9 of the PHQ-9) was included as an indicator

variable. Homoscedasticity was not assumed, and robust standard errors were computed.

Ethical Considerations

This study represents a secondary analysis of pre-existing deidentified data. The study team does not have access to the participants or to the participants' identifying information and does not intend to recontact participants. This study protocol was reviewed by the Advarra Institutional Review Board (IRB) and determined to be exempt from IRB oversight, as deidentified secondary data analysis is generally not regarded as human subjects research; this is in accordance with the US Department of Health and Human Services regulations for the protection of human subjects in research (45 CFR 46) [19].

Results

Using the baseline sample of 9165 members, Figure 2 shows the distribution of baseline scores. The scores are centered on a mean of 23.84 (SD 6.56) and a median of 24 (IQR 8) out of 40, but there is significant variance in these scores. A total of 81.0% (n=7424) had a resilience score of less than 30. Table 1 shows the mean response scores across the 10 items. Responses were highest for the questions about confidence, flexibility, and perseverance, and scores were lowest for determination and weathering emotions.

Table 2 presents the number of members and baseline score statistics for the overall sample and subgroups based on demographic characteristics and mental health outcomes at baseline. Demographic data were missing for a large portion of the sample due to irregular reporting by members' employers or health plans. Of those without missing demographic data, the majority were female and between 18 and 34 years old. Members were most likely to live in the West and South, but all four census regions were represented in the baseline sample.

For each category (gender, age, etc), a P value is presented in the category's first row testing whether the difference in mean scores across the category was statistically significant. For all categories except gender and census region, the mean baseline score was statistically different across groups at the 1% level. In the row for groups with missing data, the P value corresponds to a t test of the difference in mean baseline scores between those with and without data.

For all categories, members with missing data had significantly different scores than those without missing data. The mean scores for those with missing demographic data were lower than for those who were not missing data. While we do not have direct evidence regarding why this is the case, we can at least conclude that the resilience of members with parent organizations (ie, employers or health plans) that did not send demographic data was lower than for members with parent organizations that did report these data. The mean resilience scores for members with missing PHQ-9 and GAD-7 scores followed a different pattern in that they fell between the mean scores for clinical and subclinical members. For both the PHQ-9 and the GAD-7, the mean scores for members with missing data were considerably closer to those of subclinical members, suggesting that the set of members with missing PHQ-9 and

GAD-7 data were disproportionately subclinical, relative to the sample overall.

A significant portion of the sample screened positive for moderate to severe depression, based on the PHQ-9, or moderate to severe anxiety, based on the GAD-7. While anxiety and depression symptoms correlated with baseline scores (Table 3),

there was significant overlap in the distribution of resilience scores for members with and without clinical depression or anxiety. Figures 3 and 4 display these distributions. Notably, a significant portion of members without clinical depression (3659/4841, 75.6%) and members without clinical anxiety (3907/5130, 76.2%) had low resilience.

Figure 2. Distribution of baseline CD-RISC 10 resilience scores for 9165 members. CD-RISC 10: 10-item Connor-Davidson Resilience Scale.
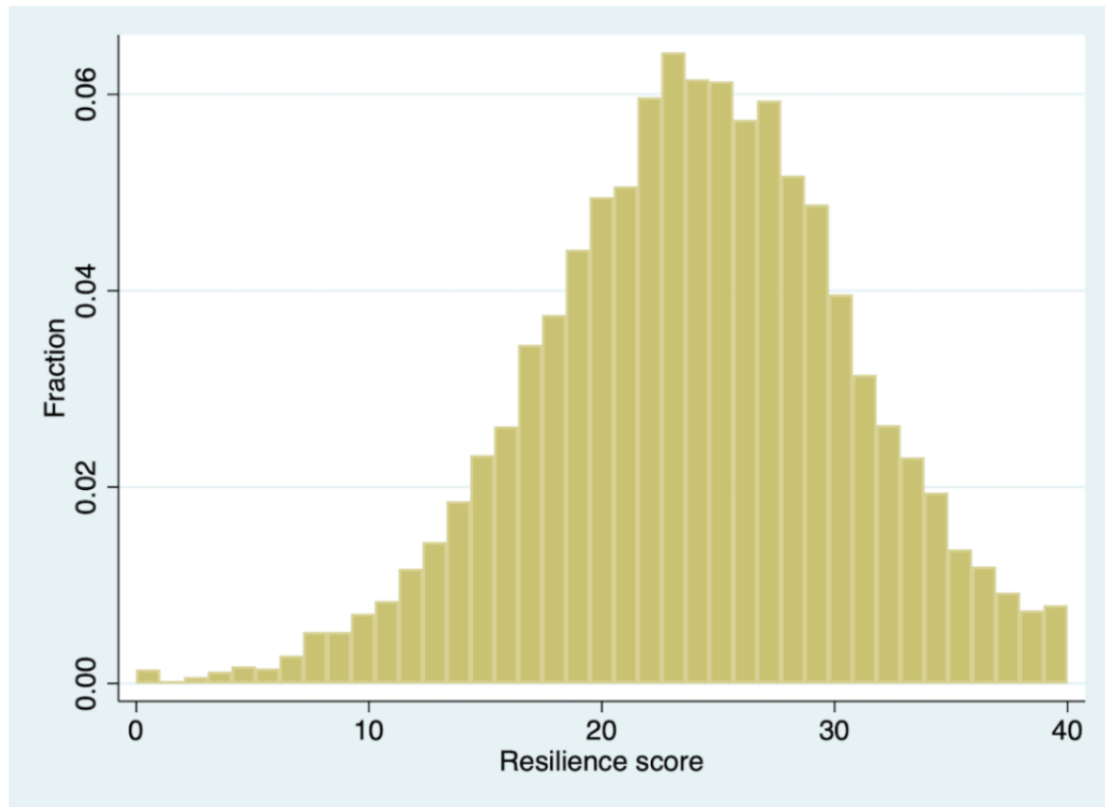
Table 1. Baseline scores for individual questions.

Question topic	CD-RISC 10[a] score, mean (SD)
Confidence	2.67 (0.95)
Determination	1.92 (1.02)
Flexibility	2.65 (0.83)
Focus	2.19 (0.99)
Grit	2.51 (1.02)
Perseverance	2.70 (0.93)
Personal growth	2.25 (0.97)
Positivity	2.36 (1.03)
Self-reliance	2.53 (0.86)
Weathering emotions	2.10 (0.99)

[a]CD-RISC 10: 10-item Connor-Davidson Resilience Scale; each item is rated on a scale from 0 (not true at all) to 4 (true nearly all the time).

Table 2. Baseline characteristics.

Variable	Participants (N=9165), n (%)	Mean (SD)	Median (IQR)	P value[a]
All participants	9165 (100)	23.84 (6.56)	24 (8)	N/A[b]
Gender				
Female	3543 (38.7)	23.91 (6.54)	24 (8)	.33
Male	1630 (17.8)	24.10 (6.83)	24 (8)	
Missing gender	3992 (43.6)	23.66 (6.47)	24 (9)	.03
Age (years)				
18-24	525 (5.7)	22.87 (6.06)	23 (8)	<.001
25-34	2368 (25.8)	24.08 (6.24)	24 (8)	
35-44	1233 (13.4)	24.46 (6.65)	25 (9)	
45-64	951 (10.4)	24.21 (6.73)	25 (9)	
≥65	48 (0.5)	23.67 (7.03)	23 (10)	
Missing age	4040 (44.1)	23.54 (6.71)	24 (9)	<.001
US Region				
West	2408 (26.3)	23.88 (6.46)	24 (8)	.39
Midwest	762 (8.3)	24.15 (6.25)	24 (8)	
South	2394 (26.1)	23.91 (6.83)	24 (9)	
Northeast	1195 (13.0)	24.22 (6.31)	25 (8)	
Missing region	2406 (26.3)	23.43 (6.61)	24 (9)	<.001
Baseline CD-RISC 10[c] resilience score				
High resilience (score ≥30)	1741 (19.0)	33.00 (2.68)	32 (4)	<.001
Low resilience (score <30)	7424 (81.0)	21.69 (5.21)	22 (7)	
Baseline PHQ-9[d] score				
Score ≥10	3641 (39.7)	21.41 (6.67)	22 (9)	<.001
Score <10	4841 (52.8)	25.47 (5.94)	26 (7)	
Missing score	683 (7.5)	25.18 (6.25)	26 (8)	<.001
Baseline GAD-7[e] score				
Score ≥10	3352 (36.6)	21.43 (6.69)	22 (9)	<.001
Score <10	5130 (56.0)	25.23 (6.04)	25 (8)	
Missing score	683 (7.5)	25.18 (6.25)	26 (8)	<.001

[a]P values were based on Welch 2-tailed t tests, which were used to analyze differences in means across subgroups; values are reported in the first row of a category with multiple subgroups. In the row for groups with missing data, the P value corresponds to a t test of the difference in mean baseline scores between those with and without data.

[b]N/A: not applicable; P values were not calculated for this variable.

[c]CD-RISC 10: 10-item Connor-Davidson Resilience Scale.

[d]PHQ-9: 9-item Patient Health Questionnaire.

[e]GAD-7: 7-item Generalized Anxiety Disorder scale.

XSL•FO
RenderX

Table 3. Ordinary least squares regression of baseline resilience scores (N=9165).

Variable	Outcome: baseline resilience score (0-40), b (SE)[a]	P value
Gender		
Female	Reference	N/A[b]
Male	0.15 (0.188)	.43
Missing gender	0.025 (0.186)	.90
Age (years)		
18-24	Reference	N/A
25-34	0.54 (0.28)	.052
35-44	0.6 (0.311)	.053
45-64	0.54 (0.323)	.09
≥65	–0.39 (1.01)	.70
Missing age	0.38 (0.269)	.16
US region		
West	Reference	N/A
Midwest	0.026 (0.254)	.92
South	0.14 (0.185)	.46
Northeast	0.077 (0.217)	.73
Missing region	–0.26 (0.222)	.24
Baseline PHQ-9[c] score		
Score ≥10	Reference	N/A
Score <10	2.3 (0.171)	<.001
Baseline GAD-7[d] score		
Score ≥10	Reference	N/A
Score <10	2 (0.168)	<.001
Suicidal ideation		
Absent	Reference	N/A
Present	–2.1 (0.206)	<.001
Missing PHQ-9 and GAD-7 scores	3.5 (0.278)	<.001
Constant	21 (0.305)	<.001

[a]The R^2 value is 0.12 and the adjusted R^2 value is 0.12.

[b]N/A: not applicable; no coefficient or P value was calculated for the reference values.

[c]PHQ-9: 9-item Patient Health Questionnaire.

[d]GAD-7: 7-item Generalized Anxiety Disorder scale.

Figure 3. Distribution of CD-RISC 10 resilience scores by PHQ-9 threshold. CD-RISC 10: 10-item Connor-Davidson Resilience Scale; PHQ-9: 9-item Patient Health Questionnaire.
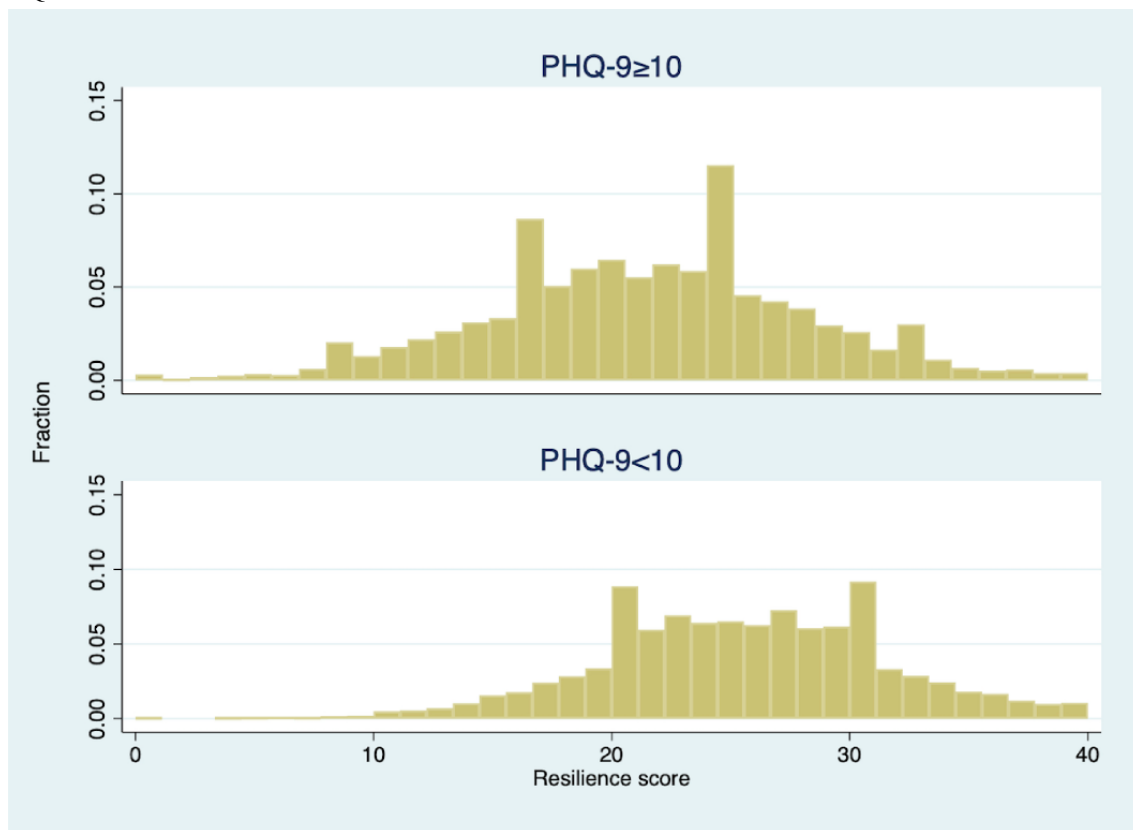
Figure 4. Distribution of CD-RISC 10 resilience scores by GAD-7 threshold. CD-RISC 10: 10-item Connor-Davidson Resilience Scale; GAD-7: 7-item Generalized Anxiety Disorder scale.
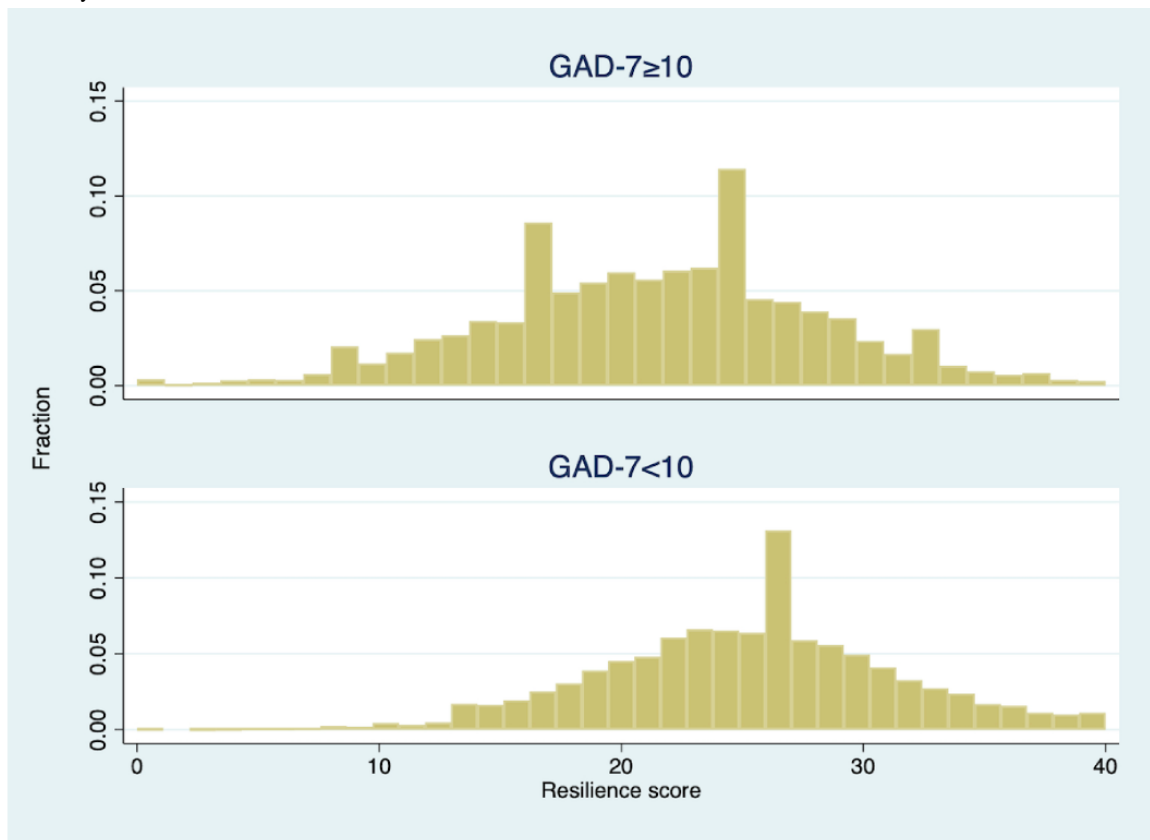
Table 3 presents results from a multivariate OLS regression using demographic and baseline mental health outcomes to predict baseline resilience scores. Point estimates and standard errors are included in the table. Holding all else constant, gender does not significantly predict scores. The youngest members (aged 18 to 24 years) had significantly lower scores than those aged 25 to 64 years and higher scores than those aged 65 years and above; the latter difference was not statistically significant. There were no statistically significant differences across census regions. Baseline scores for members with no or mild depression or anxiety were 2.3 or 2.0 points higher, respectively, than those with moderate to severe depression or anxiety. Consistent with the hypothesis that members with missing PHQ-9 and GAD-7 scores are more similar to subclinical members, the scores for these members were 4.3 points higher than for members with clinical depression (ie, the leave-out comparison group). Self-reported suicidal ideation was associated with a significant reduction of 2.1 points in baseline resilience.

Discussion

Principal Findings

This study found that individuals accessing the Ginger system had, on average, low baseline resilience levels (median CD-RISC 10 score of 24), well below prior benchmarks of the US general population and in line with studies of veterans with PTSD and depression [20]. This is not surprising given that the sample includes individuals seeking out mental health services and the data collection period, which coincided with spikes of COVID-19 cases and other disruptive world events. A significant portion of the sample screened positive at baseline for moderate to severe depression (3641/8482, 42.9%, based on the PHQ-9) and moderate to severe anxiety (3352/8482, 39.5%, based on the GAD-7). A total of 81.0% of members had a baseline resilience score of less than 30.

Our descriptive analysis shows that younger members tended to have lower resilience, which is consistent with several studies of adolescents and young adults in the United States [3,21]. Baseline resilience scores for members with no or mild depression or anxiety were higher than those for members with moderate to severe depression or anxiety. This association was even stronger for members who reported suicidal ideation. This is consistent with findings from other studies showing that self-reported mental health diagnoses were negatively associated with higher resilience, and adults with reported low or normal levels of resilience were more likely to experience mental distress compared to those with high resilience [11,21]. However, it is interesting to note that despite having relatively higher resilience scores, members with no or mild depression or anxiety still had low resilience scores on average. This highlights the need for mental health support among individuals who might not typically be recommended for treatment based on traditional clinical assessments, like the PHQ-9 and the GAD-7.

Strengths and Limitations

A major strength of this study was the large number of participants; this is one of the largest studies using the CD-RISC 10 measure, which allows for certain subgroup analyses. This is also one of the first applications of this measure in a large-scale real-world setting, in contrast to smaller controlled research settings. Incorporating a strength-based measure like resilience, in contrast to symptom-based measures like the PHQ-9 and the GAD-7, allows us to better understand the needs of individuals seeking mental health services.

There are several limitations to this study. Of the 9165 Ginger participants in the study, 43.6% had missing data for gender, and 44.1% had missing data for age. Typically, these demographic data are shared by members' organizations through which they access Ginger services (eg, employers). Some organizations do not share demographic data. Given the amount of missing data in our sample, we acknowledge the need for further research that focuses on the relationship between resilience and demographics. In part due to incomplete data reporting by parent organizations, Ginger has launched the capability for members to elect to self-report their demographic information in the app. These self-reported data will supplement the parent organization–reported data and will be available for future research projects. This functionality had not been launched by the time of this analysis. For the purposes of this study, we presented results for members with missing data and we controlled for whether a member's demographic data were missing in any regression analyses. However, because the Ginger platform is offered through employers, the survey respondents were working-age adults, which suggests that these findings may generalize to the professional workforce and those enrolled in health benefits through their employer. Further, the current data did not include information on other sociodemographic or contextual factors (eg, marital status, family composition, significant life events, sources of social support, and educational level) that might be related to resilience and mental health, and that may have been of particular significance during the pandemic.

Future Research

Given that behavioral health providers often focus on clinical symptoms, such as those measured by the PHQ-9 and the GAD-7, a deeper understanding of nonclinical outcomes, such as resilience, is increasingly important to the growing digital behavioral health industry. This is especially true given that many members seeking behavioral health support do not experience clinical symptoms. For example, the majority of members in this study screened negative for depression and anxiety, signaling a need to not only track other outcomes that are associated with members' well-being but also to understand which specific interventions have an impact on these outcomes, the expected size of these impacts, and which subpopulations may respond differently. This study points to many directions for future research, primarily looking at how these scores evolve over time, with a particular focus on whether the interaction of resilience and clinical symptoms (eg, depression and anxiety) impacts members' responses to behavioral health interventions. Given that this data set included PHQ-9 and GAD-7 scores, future studies could examine the relationship between resilience and symptoms of anxiety and depression, for example, whether increased baseline resilience is associated with a higher likelihood of symptom improvement and faster time to improvement (ie, less use of care services). Additionally, future

research could look at more detailed classifications of use and conversational features extracted via natural language processing of text messages to better understand which factors have stronger associations with increased resilience.

Conclusions

Resilience is a construct often referred to but less often defined and measured, particularly in clinical settings that tend to focus more on symptom-based measures. In this study, we found that members accessing mental health services from January to August 2021 had extremely low baseline resilience, in line with prior studies of trauma survivors, which highlights the need for expanding access to care. Overall, younger members and those with higher levels of depression and anxiety at intake reported lower levels of resilience at baseline. Notably, members with no or mild depression or anxiety still had low resilience scores on average, demonstrating the need for mental health support among individuals without clinical symptoms. Future research will examine changes in resilience over time in addition to factors associated with those changes.

Conflicts of Interest

GG, SK, and ES are current paid employees of Ginger (Headspace Health).

References

1. Connor KM, Davidson JRT. Development of a new resilience scale: The Connor-Davidson Resilience Scale (CD-RISC). Depress Anxiety 2003;18(2):76-82. [doi: 10.1002/da.10113] [Medline: 12964174]
2. Hu T, Zhang D, Wang J. A meta-analysis of the trait resilience and mental health. Pers Individ Dif 2015 Apr;76:18-27. [doi: 10.1016/j.paid.2014.11.039]
3. Campbell-Sills L, Cohan SL, Stein MB. Relationship of resilience to personality, coping, and psychiatric symptoms in young adults. Behav Res Ther 2006 Apr;44(4):585-599. [doi: 10.1016/j.brat.2005.05.001] [Medline: 15998508]
4. Zadok-Gurman T, Jakobovich R, Dvash E, Zafrani K, Rolnik B, Ganz AB, et al. Effect of inquiry-based stress reduction (IBSR) intervention on well-being, resilience and burnout of teachers during the COVID-19 pandemic. Int J Environ Res Public Health 2021 Apr 01;18(7):3689 [FREE Full text] [doi: 10.3390/ijerph18073689] [Medline: 33916258]
5. Waller MA. Resilience in ecosystemic context: Evolution of the concept. Am J Orthopsychiatry 2001 Jul;71(3):290-297. [doi: 10.1037/0002-9432.71.3.290] [Medline: 11495331]
6. Enrique A, Mooney O, Salamanca-Sanabria A, Lee C, Farrell S, Richards D. Assessing the efficacy and acceptability of an internet-delivered intervention for resilience among college students: A pilot randomised control trial protocol. Internet Interv 2019 Sep;17:100254. [doi: 10.1016/j.invent.2019.100254]
7. Smith B, Shatté A, Perlman A, Siers M, Lynch W. Improvements in resilience, stress, and somatic symptoms following online resilience training: A dose-response effect. J Occup Environ Med 2018 Jan;60(1):1-5 [FREE Full text] [doi: 10.1097/JOM.0000000000001142] [Medline: 28820863]
8. Ciro D, Pietrzak RH, Lee RJ, Rodriguez J, Singh R, Salim R, et al. Acculturation, coping, and PTSD in Hispanic 9/11 rescue and recovery workers. Psychol Trauma 2021 Jan;13(1):84-93. [doi: 10.1037/tra0000624] [Medline: 33382331]
9. Feingold JH, Hurtado A, Feder A, Peccoralo L, Southwick SM, Ripp J, et al. Posttraumatic growth among health care workers on the frontlines of the COVID-19 pandemic. J Affect Disord 2022 Jan 01;296:35-40 [FREE Full text] [doi: 10.1016/j.jad.2021.09.032] [Medline: 34587547]
10. Feder A, Fred-Torres S, Southwick SM, Charney DS. The biology of human resilience: Opportunities for enhancing resilience across the life span. Biol Psychiatry 2019 Sep 15;86(6):443-453. [doi: 10.1016/j.biopsych.2019.07.012] [Medline: 31466561]
11. Riehm KE, Brenneke SG, Adams LB, Gilan D, Lieb K, Kunzler AM, et al. Association between psychological resilience and changes in mental distress during the COVID-19 pandemic. J Affect Disord 2021 Mar 01;282:381-385 [FREE Full text] [doi: 10.1016/j.jad.2020.12.071] [Medline: 33421866]
12. Killgore WDS, Taylor EC, Cloonan SA, Dailey NS. Psychological resilience during the COVID-19 lockdown. Psychiatry Res 2020 Sep;291:113216 [FREE Full text] [doi: 10.1016/j.psychres.2020.113216] [Medline: 32544705]
13. Golden EA, Zweig M, Danieletto M, Landell K, Nadkarni G, Bottinger E, et al. A resilience-building app to support the mental health of health care workers in the COVID-19 era: Design process, distribution, and evaluation. JMIR Form Res 2021 May 05;5(5):e26590 [FREE Full text] [doi: 10.2196/26590] [Medline: 33872189]
14. Kaye-Kauderer H, Feingold J, Feder A, Southwick S, Charney D. Resilience in the age of COVID-19. BJPsych Adv 2021 Feb 17;27(3):166-178 [FREE Full text] [doi: 10.1192/bja.2021.5]
15. Kunkle S, Yip M, Hunt J, Ξ W, Udall D, Arean P, et al. Association between care utilization and anxiety outcomes in an on-demand mental health system: Retrospective observational study. JMIR Form Res 2021 Jan 26;5(1):e24662 [FREE Full text] [doi: 10.2196/24662] [Medline: 33496679]

XSL•FO
RenderX

16. Kunkle S, Yip M, Ξ W, Hunt J. Evaluation of an on-demand mental health system for depression symptoms: Retrospective observational study. J Med Internet Res 2020 Jun 18;22(6):e17902 [FREE Full text] [doi: 10.2196/17902] [Medline: 32554387]

17. Campbell-Sills L, Forde DR, Stein MB. Demographic and childhood environmental predictors of resilience in a community sample. J Psychiatr Res 2009 Aug;43(12):1007-1012. [doi: 10.1016/j.jpsychires.2009.01.013] [Medline: 19264325]

18. Vaishnavi S, Connor K, Davidson JRT. An abbreviated version of the Connor-Davidson Resilience Scale (CD-RISC), the CD-RISC2: Psychometric properties and applications in psychopharmacological trials. Psychiatry Res 2007 Aug 30;152(2-3):293-297 [FREE Full text] [doi: 10.1016/j.psychres.2007.01.006] [Medline: 17459488]

19. 45 CFR 46. US Department of Health & Human Services, Office for Human Research Protections. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html [accessed 2022-03-21]

20. Wingo AP, Briscione M, Norrholm SD, Jovanovic T, McCullough SA, Skelton K, et al. Psychological resilience is associated with more intact social functioning in veterans with post-traumatic stress disorder and depression. Psychiatry Res 2017 Mar;249:206-211. [doi: 10.1016/j.psychres.2017.01.022] [Medline: 28119173]

21. Hartley M. Assessing and promoting resilience: An additional tool to address the increasing number of college students with psychological problems. J Coll Couns 2012 Apr 11:37-51. [doi: 10.1002/j.2161-1882.2012.00004.x]

Abbreviations

CD-RISC: Connor-Davidson Resilience Scale
CD-RISC 10: 10-item Connor-Davidson Resilience Scale
GAD-7: 7-item Generalized Anxiety Disorder scale
IRB: Institutional Review Board
OLS: ordinary least squares
PHQ-9: 9-item Patient Health Questionnaire
PTSD: posttraumatic stress disorder

Original Paper

App-Based Ecological Momentary Assessment to Enhance Clinical Care for Postpartum Depression: Pilot Acceptability Study

Holly Krohn[1], MPH; Jerry Guintivano[1], PhD; Rachel Frische[1], MPH, MBA, MD; Jamie Steed[1], BSc; Hannah Rackers[1], MPH; Samantha Meltzer-Brody[1], MPH, MD

Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Holly Krohn, MPH
Department of Psychiatry
University of North Carolina at Chapel Hill
333 S. Columbia Street
MacNider 304 CB 7160
Chapel Hill, NC, 27599
United States
Phone: 1 919 445 0218
Email: holly_krohn@med.unc.edu

Abstract

Background: Wearable tracking devices and mobile health technology are increasingly used in an effort to enhance clinical care and the delivery of personalized medical treatment. Postpartum depression is the most frequently diagnosed complication of childbirth; however, significant gaps in screening and treatment remain.

Objective: This study aims to investigate the clinical utility, predictive ability, and acceptability of using ecological momentary assessment to collect daily mood, sleep, and activity data through the use of an Apple Watch and mobile app among women with postpartum depression.

Methods: This was a pilot study consisting of 3 in-person research visits over the course of a 6-week enrollment period. Questionnaires to assess depression, anxiety, and maternal functioning were periodically collected, along with daily self-reported symptoms and passively collected physiological data via an Apple Watch. Feedback was collected from study participants and the study clinician to determine the utility and acceptability of daily tracking. Logistic regression was used to determine whether mood scores in the 2 weeks before a visit predicted scores at follow-up. Compliance with daily assessments was also measured.

Results: Of the 26 women enrolled, 23 (88%) completed the 6-week study period. On average, the participants completed 67% (34.4/51.5 days) of all active daily assessments and 74% (38/51.5 days) of all passive measures. Furthermore, all 23 participants completed the 3 required visits with the research team. Predictive correlations were found between self-reported mood and Edinburgh Postnatal Depression Scale score at follow-up, self-reported anxiety and EDPS, and sleep quality and Edinburgh Postnatal Depression Scale.

Conclusions: Using ecological momentary assessment to track daily symptoms of postpartum depression using a wearable device was largely endorsed as acceptable and clinically useful by participants and the study clinician and could be an innovative solution to increase care access during the COVID-19 pandemic.

Introduction

Background

Postpartum depression (PPD) is a common yet serious disorder of the perinatal period. Despite being the most frequently diagnosed complication of childbirth, there are pervasive gaps in the screening, detection, and treatment of women with this debilitating condition [1]. Lifetime positive screening prevalence estimates for PPD range from 10% to 15% [2-4]. If inadequately treated, the condition can lead to devastating consequences for

families; PPD is considered the greatest risk factor for maternal suicide and infanticide [5]. Research has shown a multitude of harmful effects to the mother-baby dyad, including higher medical costs [6], earlier discontinuation of breastfeeding [7], negative impact on the baby's brain development [8,9], and an increased risk for long-term mental health problems in children of women with PPD [9].

Complex psychiatric disorders, such as PPD, require thoughtful clinical tools for decision support, which are an increasingly important component of providing high-quality health care. Vital information pertaining to a patient's condition can be derived from ecological momentary assessment (EMA) data. EMA includes various methods for data collection, traditionally via paper-and-pencil questionnaires or diaries and more recently via mobile phone or device reporting at certain designated time points throughout the day [10]. This method of *real-world* data collection has been shown to reduce recall bias in reporting mood and to highlight dynamic changes a patient may experience day to day and can be easily integrated with physiological data for a broader picture of a patient's experience outside of the clinical setting [11,12]. EMA ratings can be *prompted* or passively collected, especially in the context of smartphones and smart technology with integrated passive data collection such as heart rate and activity that are constantly sampled throughout the day. In recent years, EMA data have been shown to have the potential to predict oncoming mood and anxiety episodes [12-16]. A review of the literature did not identify any other studies examining mood tracking using EMA in a population with PPD, although recent studies have examined substance use disorder, posttraumatic stress disorder, racial disparities, and postpartum health in perinatal populations using EMA [17,18]. EMA has been used to understand the course of affective experience during pregnancy but not postpartum [19]. Given the significant health burden associated with PPD, there is a need for innovative technology that can enhance and personalize the clinical care of perinatal women at risk for mood disorders. The COVID-19 pandemic has led to devastating consequences for perinatal women, making innovative care solutions all the more relevant [20]. With the closing of medical offices and schools and the decrease in social and community support, accessible and telemedicine-based care options are critical to ensure patient safety and well-being [21]. To address this problem, our team developed and pilot-tested an app-based EMA module within an existing research app for women with PPD to measure mood, anxiety, sleep, and activity in women experiencing PPD.

The existing research app, *PPD ACT* (now *Mom Genes Fight PPD*), was released in 2016 to rapidly and efficiently recruit, consent, screen and enable DNA collection from women with a lifetime history of PPD [22]. Interested participants downloaded the app, consented to the study, completed a series of PPD screening questionnaires, and submitted a saliva sample if eligible. In an effort to increase engagement with the PPD ACT app, the *PPD ACT Apple Watch Module* was created to allow for daily tracking of mood, anxiety, sleep quality, and exercise (via heart rate and step tracking) for participants to better understand how each activity can affect symptoms of

depression and anxiety and to be able to easily share these data with a clinician.

Objectives

It has been well documented that women experiencing PPD face a number of barriers to obtaining appropriate medical care. Prominent barriers include stigma-related reasons such as a hesitancy to admit to experiencing depression and a tendency to minimize symptoms [23]. Engaging patients in care through the use of device-based apps and EMA is a potentially powerful tool to minimize treatment barriers while providing a rich set of data for clinician use to inform more personalized patient care. The primary goal of this pilot study is to determine the clinical utility and predictive ability of using an app-based daily tracking tool to enhance PPD treatment. We also assessed patient compliance with the protocol and feedback on the experience and usefulness of the EMA data by both participants and the study clinician.

Methods

Participants

Recruitment of English-speaking women aged >18 years and currently experiencing PPD occurred from February 2017 to March 2018 in specialized perinatal psychiatry clinics at UNC Hospitals in Chapel Hill, North Carolina, and the surrounding community. Participants were screened for PPD using the Edinburgh Postnatal Depression Scale (EPDS) [24] or identified through medical records. Those who were <7 months postpartum, with an EPDS score of ≥13 or a positive current PPD diagnosis in the medical record were approached to participate in the study. Case status was confirmed within *PPD ACT* using the EPDS lifetime version [25]. The study was additionally advertised in the community via flyers and targeted Facebook advertising. Enrollment was also limited to women who owned an Apple iPhone 5 or a newer model to pair with the Apple Watch.

Ethics Approval

This study was approved by the University of North Carolina Institutional Review Board Office of Human Research Ethics (number 15-2165). All participants provided written informed consent and signed the Health Insurance Portability and Accountability Act.

PPD ACT Apple Watch Module

To collect EMA data via the *PPD ACT* app, we created the *PPD ACT Apple Watch Module* with the support of app development firm Little Green Software using Apple's CareKit framework, a tool allowing for app-based data collection aimed at understanding and management of health conditions. The Apple Watch Module allows women to log their self-reported mood, anxiety, sleep quality, and medication use (if applicable) daily, while using an Apple Watch to passively collect physiological data on their daily activity (steps taken), heart rate (periodic beats per minute), and sleep. The Sleep++ app, a third-party app, was downloaded on the Apple Watch and used to passively log nightly sleep duration. Integrated sleep tracking was not yet available on the Apple Watch, and a third-party app was used

instead. Although categorized as passive, the Sleep++ app did require participant input to log when they went to sleep and woke up each day. Mood, sleep, and physiological data were shared with the study clinician via the participant's PPD ACT app at research visits to provide critical information about the daily experience of PPD outside of the clinical evaluation.

Equipment

A total of 24 Series 2 Apple Watches (42 mm) were donated for study use by Apple, Inc. The Apple Watches were then loaned to participants for the 6-week duration of the study to be used for data collection. Participants signed an Apple Watch Loan Agreement to outline return of the Apple Watch at the end of study participation. Participants were required to own an Apple iPhone 5 or a newer model to pair with the Apple Watch for functionality. The phone was also needed to download the *PPD ACT* app, which participants concurrently enrolled in. The specialized Apple Watch Module was available on the *PPD ACT* app via an enrollment code provided by the research team.

Assessment Measures

Overview

Participants were instructed to wear the Apple Watch continuously for the 6-week enrollment period. Each day, participants were asked to open the *PPD ACT* app Apple Watch Module and fill out a 10-point Likert scale pertaining to their daily mood, anxiety, sleep quality (1 being the worst ever experienced and 10 being the best ever experienced), and medication use (if applicable). A full list of assessments is provided in Multimedia Appendix 1. Participants were asked to rate their sleep quality in the app daily upon waking and to complete the mood and anxiety assessments at the end of each day. Figure 1 shows examples of in-app questionnaires. In addition, the Apple Watch passively collected daily activity and heart rate. Sleep was *semipassively* tracked using the Sleep++ app, which required the participant to open the app before going to bed to indicate the start of the sleep period and upon waking to indicate the stop of the sleep period. Participants were able to track their data over the course of the study via summary charts provided in the Apple Watch Module.

Participants were asked to attend 3 research visits at the UNC Hospitals Perinatal Psychiatry clinic over the course of 6 weeks. At these visits, they met with research coordinators (HK and JS) and the research clinician and completed self-reported questionnaires including the Patient Health Questionnaire-9 (PHQ-9), the Generalized Anxiety Disorder 7-item scale (GAD-7), the EPDS, and the Barkin Index of Maternal Functioning (BIMF). The PHQ-9 and EPDS are commonly administered together in research studies of PPD and have been shown to be reliable and valid measures for identifying major depressive episodes in the general population (and in populations with PPD for the EPDS) [26,27]. As the scales measure slightly different constructs, it is beneficial to administer both. Studies have found discordance to occur between PHQ-9 and EPDS scores [28], and we also had 5 instances where the EPDS and PHQ-9 were discordant over the 3 visits. However, as the scores were not being used diagnostically, we did not find this to be an issue. Participants also completed a feedback survey at their second and third visits about their experience using the Apple Watch. After visit 3, the research clinician completed a feedback questionnaire for each participant on the clinical utility of the EMA measures.

Figure 1. Apple Watch Module questionnaire screens.
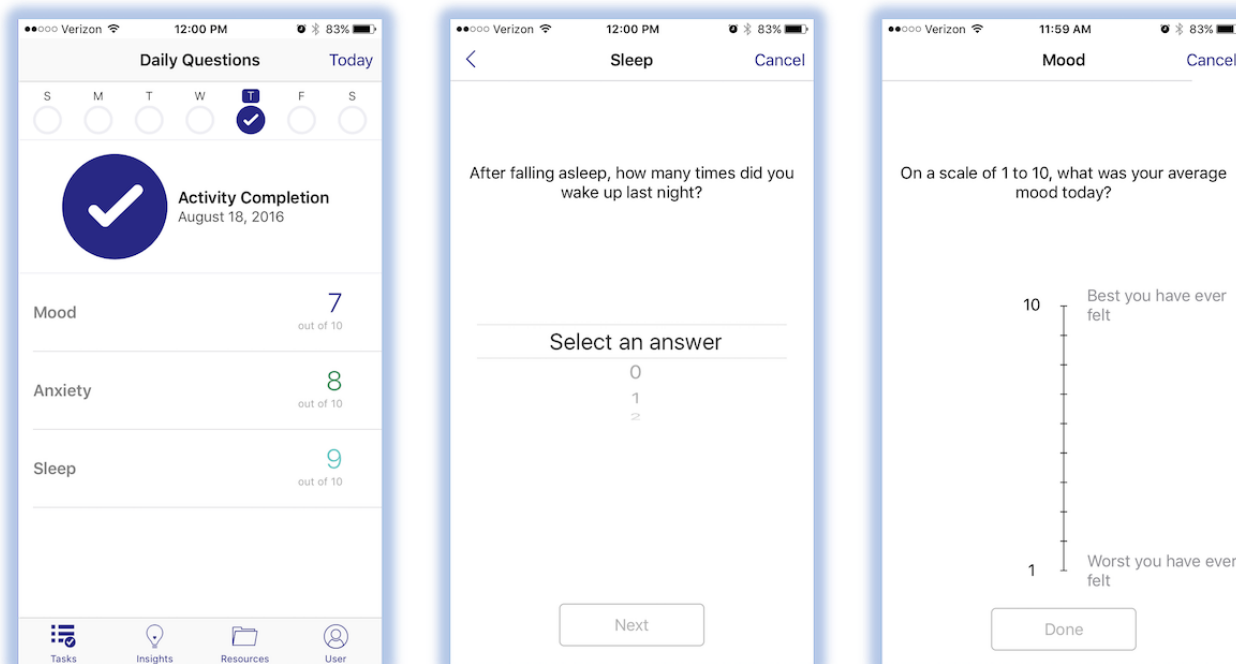
Instruments

PHQ-9 Assessment

The PHQ-9 is a validated [29], self-administered questionnaire commonly used in clinical settings to guide screening or measure recovery and treatment response. This 9-item questionnaire was originally designed to measure the presence or severity of depressive symptoms. Patients are asked to rate items based on their experience over the past 2 weeks. Each item is rated on a Likert scale ranging from 0 to 3 and scored based on the frequency of symptoms and functional impairment. Four anchors are included (0=not at all, 1=several days, 2=more than half the days, and 3=nearly every day). Ratings produce a summary score between 0 and 27. A score ≤4 indicates no depressive symptoms, 5 to 9 mild depression, 10 to 14 moderate depression, 15 to 19 moderately severe depression, and ≥20 severe depression [30].

GAD-7 Assessment

The GAD-7 is a well-validated [31] clinical tool for measuring generalized anxiety disorder and the three most common anxiety disorders (panic disorder, posttraumatic stress disorder, and social anxiety disorder) [30]. The GAD-7 assesses the frequency and severity of anxiety symptoms in the past 2 weeks using the same Likert scale and anchors as those used in the PHQ-9. Scores of 5, 10, and 15 are taken as the cut-off points for mild, moderate, and severe anxiety, respectively. Using the threshold score of 10, the GAD-7 has a sensitivity of 89% and a specificity of 82% for generalized anxiety disorder [32].

EPDS Assessment

The EPDS is a 10-item self-reported questionnaire that has been well validated [33] and shown to be reliable and sensitive in identifying women at risk for PPD [24]. Women are asked to report on symptoms over the previous 7 days. Response categories are scored on a 4-point scale from 0 to 3 according to the increased severity of the symptom. A total score of ≥13 indicates the likelihood of depression and suggests the need for follow-up care [24].

BIMF Assessment

The BIMF was developed to measure the maternal functional status of women in the 12 months following childbirth [34]. The BIMF is a self-reported, 20-item questionnaire designed to address seven domains of maternal functioning, namely, (1) self-care, (2) infant care, (3) mother and child interaction, (4) social support, (5) psychological well-being, (6) management, and (7) adjustment [35]. The range of scores are from 0 to 120, with a total score of 120 indicating optimal maternal functioning. The BIMF has been validated in populations with PPD [34].

Data Analysis

Qualitative analyses were conducted using R (version 3.6.3; R Foundation for Statistical Computing). Descriptive statistics are reported using percentages for categorical variables and means (SDs) for continuous variables. To examine the predictive ability of self-report ratings for mood, anxiety, and sleep on outcomes at clinical follow-up visits, linear regression models were created for each participant and each self-report rating. Using self-report data for the 2 weeks before a follow-up visit, linear models were constructed as rating ~ day + visit, where rating is the daily rating, day is the numerical day before the follow-up visit (range 1-14), and visit indicates which visit the self-report data precedes (visit 2 or visit 3). From these models, we predicted the given self-report measure on the given follow-up visit. We measured the predictive ability of these models using root-mean-square error (RMSE). Pearson correlations were performed to measure the strength of the association between predicted self-report values and EPDS scores and actual self-report values on the follow-up visit days. *P* values reported are those from Pearson correlation tests.

Patient and Provider Feedback

Feedback surveys were developed specifically for this study to assess the individual tracking components of the module and the comfort and ease of wearing the Apple Watch. The participants' surveys incorporated free-text questions to assess technical issues, ease, and value of the intervention. The provider survey incorporated free-text questions to evaluate clinical utility and perceived participant value provided to users at home and in a clinical setting. Please see Multimedia Appendix 2 to view the feedback surveys. Feedback responses were evaluated as qualitative data [36]. We utilized a thematic analysis approach [37], and a single coder (HR) tracked developing codes and emerging preliminary themes during analysis and organized these codes into final themes. The selection of codes and original responses were reviewed by a second team member (JS) to establish consensus and increase validity and reliability.

Results

Participants

A total of 26 women were enrolled in the study, but only 23 completed the study. The 3 women who withdrew initially enrolled but did not initiate study procedures before withdrawal, citing lack of time to complete the study as the reason for leaving. Descriptive statistics are presented in Table 1.

Data were tracked across four domains in the study: (1) participant compliance with EMA and study measures (ie, how often did participants complete the required daily activities and how often did participants wear the Apple Watch); (2) assessments measuring depression, anxiety, and maternal functioning (PHQ-9, GAD-7, EPDS, and BIMF) across study enrollment; (3) participant feedback on utility and experience of wearing the Apple Watch and completing tracking measures; and (4) clinician feedback on the utility of tracked data. Of the 23 participants who completed the study, only 21 (91%) completed all data points for each of the 3 visits, whereas 2 (9%) participants did not enter data for all of the assessment measures at visits 2 and 3; those missing data were not included in the results of Table 2.

Table 1. Demographics (n=23).

Study variable	Values
Age (years), mean (SD)	33 (7.5)
Months postpartum, mean (SD)	4.5 (1.6)
Parity, n (%)	
Primiparous	11 (48)
Multiparous	12 (52)
PPD[a] episodes, n (%)	
First-episode PPD	16 (70)
Multiple episodes	7 (30)
Marital status, n (%)	
Married	18 (78)
Divorced	2 (9)
Never married	3 (13)
Medication use, n (%)	
Currently prescribed medication	16 (70)
No medication	7 (30)
Race and ethnicity, n (%)	
Asian	2 (9)
Black or African American	1 (4)
Hispanic or Latina	2 (9)
White	18 (78)
Breastfeeding, n (%)	
Yes	11 (48)
No	12 (52)

[a]PPD: postpartum depression.

Table 2. Depression, anxiety, and maternal functioning assessment scores.

Assessment measures	Visit 1, mean (SD)	Visit 2, mean (SD)	Final visit, mean (SD)
PHQ-9[a]	12 (5)	10 (5)	7 (5)
GAD-7[b]	11 (5)	10 (5)	7 (5)
EPDS[c]	14 (5)	12 (5)	11 (6)
BIMF[d]	76 (18)	83 (14)	88 (13)

[a]PHQ-9: Patient Health Questionnaire-9.

[b]GAD-7: Generalized Anxiety Disorder-7.

[c]EPDS: Edinburgh Postnatal Depression Scale.

[d]BIMF: Barkin Index of Maternal Functioning.

EMA and Compliance

The mean number of days the participants were enrolled was 52 (SD 12) days. The mean number of days enrolled was greater than the anticipated 6 weeks of enrollment (approximately 42 days), which can be accounted for by patients often being unable to come exactly 3 weeks apart, or piggybacking research appointments on regularly scheduled appointments in clinics, which are often 4 weeks apart. The compliance of measures being reported varied across the self-report and passive measure domains, with steps (activity) having the highest number of days reported (mean 45.2) and the highest percentage of total participation at nearly 90%, as illustrated in Table 3. We did find low compliance for sleep tracking, as participants indicated many technical issues with the Sleep++ app, including it not tracking sleep accurately after following app directions and discomfort wearing the watch overnight.

XSL•FO
RenderX

Table 3. Ecologic momentary assessment and compliance (mean number of days enrolled 52, SD 12).

	Days reported, mean (SD)	Completion proportion (%), mean (SD)
Active measures	33 (15)	67 (30)
Self-reported mood	32 (15)	66 (30)
Self-reported anxiety	33 (16)	68 (31)
Self-reported sleep quality	33 (15)	67 (31)
Passive measures	37 (17)	74 (28)
Heart rate	37 (13)	76 (27)
Steps	45 (13)	90 (20)
Sleep	25 (20)	47 (37)

Depression, Anxiety, and Maternal Functioning Measures and Predictive Value

Assessment measures for depression, anxiety, and maternal functioning were tracked across three research visits using the PHQ-9, GAD-7, EPDS, and BIMF. There was an overall improvement in scores between baseline and end of study assessment scores. However, all but 1 participant was in concurrent treatment for PPD, so these improvements may be accounted for by treatment attendance.

Using the collected prospective data, we attempted to predict outcomes observed at the second and third clinical visits. Data from the 2 weeks before the follow-up visit were used to train a simple regression model to predict various outcomes at follow-up. Self-reported mood on the day of follow-up was negatively associated with EPDS score at follow-up ($r=-0.49$; $P=.02$). We also found that self-reported mood in the 2 weeks before the follow-up visit predicts the self-reported mood rating on the day of follow-up (RMSE=1.21; $r=0.64$; $P<.001$). Furthermore, participants with PPD at follow-up (EPDS>12) had a higher predicted mood rating compared with those without PPD (6.4 vs 6.0; $P=.38$). In addition, self-reported anxiety was positively associated with EPDS score at follow-up ($r=0.56$; $P=.004$). Self-reported anxiety in the 2 weeks before the follow-up visit predicts the self-reported anxiety rating on the day of follow-up (RMSE=1.37; $r=0.75$; $P<.001$). Participants with PPD at follow-up (EPDS>12) had a lower predicted anxiety rating compared with those without PPD (5.2 vs 5.9; $P=.25$). In addition, self-reported sleep quality was negatively associated with EPDS score at follow-up ($r=-0.10$; $P=.66$). Self-reported sleep quality in the 2 weeks before the follow-up visit predicts the self-reported sleep quality rating on the day of follow-up (RMSE=2.2; $r=0.22$; $P=.33$), although not as strongly as predictors for mood and anxiety. Interestingly, participants with PPD at follow-up (EPDS>12) had a higher predicted sleep quality rating compared with those without PPD (5.9 vs 5.2; $P=.35$).

Clinician Feedback

Overview

Three primary themes emerged from clinician feedback regarding the utility of the data collected: (1) increased insight of participants into mental health status, (2) enhancement of patient engagement and discussion of treatment, and (3) improved monitoring of treatment effectiveness.

Increased Insight Into Mental Health Status

The watch's ability to track data including sleep patterns, heart rate, and steps along with daily logs of mood, anxiety, sleep quality, and medication increased the participants' abilities to connect patterns in these factors to their mood and anxiety symptoms. For some women, they had not understood the severity of the symptoms they were experiencing. These data gave participants and providers a tangible picture of symptoms that increased insight into mental health status. For other participants, these data provided insight into how mental health affected other behaviors. There was 1 participant who connected her activity levels with the severity of her symptoms. This participant then made efforts to reach out for more social support to increase feelings of security when leaving home with the baby.

Enhancement of Patient Engagement

The most commonly identified theme among clinician response was that the app data provided a basis for discussion and increased engagement with the participant. In addition to providing a platform for engagement in psychiatric treatment, data from the Apple Watch also prompted participants to reach out to their primary care providers regarding health concerns.

Improved Monitoring of Treatment Effectiveness

The watch data enriched observations around the effects of treatment. For those who were being treated with medication, changes in medication could be monitored using the daily logs.

Participant Feedback

Overview

Participants were amenable to use of this technology with 61% (14/23) endorsing that they would use this technology to inform personal habits, 43% (10/23) endorsing that they would continue to use this technology in medical care, and 57% (13/23) recommending this technology to others. Open-ended feedback was solicited from participants surrounding challenges and benefits to wearing the watch and tracking. These are discussed in the following sections.

Challenges of Use

The primary challenge cited was technical difficulties with the Sleep++ app (9/23, 39% participants). Participants indicated that the app did not always technically perform (would not track sleep when used as directed). These difficulties made some women doubt the accuracy of sleep measures. The second challenge noted was difficulty wearing the watch overnight (8/23, 35% participants). Specifically, the watch lighting up at night was reported as problematic, sound alert notifications at night were disturbing, and there was general discomfort associated with wearing the watch overnight due to bulk and sensitivity with wearing any jewelry or watches while sleeping. Finally, of the 23 women, 2 (9%) reported feeling more stressed and an increase in anxiety from the tracking and extra activities required each day, and 2 (9%) also reported the watchband causing a rash and being generally uncomfortable.

Benefits of Use

The primary benefit of use cited was increased insight into sleep patterns, even despite the technical difficulties with the Sleep++ app (9/23, 39% participants). Primarily, women did not realize how little sleep or how disrupted the sleep they were getting was. These data were then shared with their work, partner, or others to help improve sleep quality or length. There was 1 participant who was able to correlate her daily mood with sleep quality. Second, women reported benefit to seeing their daily or weekly activity, mostly seeing the days they did not get many steps in and a desire to increase activity on those days. Participants also enjoyed the Apple Watch itself, and 3 people subsequently purchased their own.

Discussion

Principal Findings

As a pilot trial, the primary motivation for this study was to determine if EMA has utility in the clinical care of patients with PPD by assessing both provider and patient perspectives. Overall, most study participants and the study clinician endorsed a clinical benefit from using this intervention. We were also able to demonstrate that daily self-reported mood and anxiety scores correlated with EPDS scores at study visits, which is a common and standard screening tool for PPD [3,38,39]. This knowledge could be very useful for predicting when a patient is experiencing an upswing or downswing in symptoms and for prompt early intervention. Participant compliance to both active and passive measures was encouraging, with average compliance rates of 67% for all active daily assessments and 74% of all passive measures. Study participants also identified areas for improvement, specifically about the sleep tracking app Sleep++, which was inconsistent in performance.

A review of the literature did not identify any other studies examining mood tracking using EMA in a population with PPD, although recent studies have examined substance use disorder, posttraumatic stress disorder, racial disparities, and postpartum health in perinatal populations using EMA [17,18]. However, several studies have examined compliance with tracking measures in other psychiatric domains. In a compliance study of computerized ambulatory monitoring in psychiatry, a sample of 45 participants with an anxiety disorder were found to have a 73% compliance rate with assessments over a 1-week period [40]. In a pilot study of mood ratings captured via mobile phones among participants with bipolar disorder, average compliance was 42% over a 12-week period [41], albeit with an SD of 26.6% and a range of 45.8% to 93%. Another mobile health study of EMA and mood symptoms in participants with traumatic brain injury found that the average compliance was 73.4% over an 8-week period [42]. Given these examples, the average compliance for daily mood, anxiety, and sleep quality are encouraging in a population of postpartum women juggling the daily care of an infant (many with multiple children) while experiencing depression.

For the passively collected data measures, we found even higher participation rates, particularly for heart rate and activity (steps), with compliance rates of 76% (SD 27%) and 90% (SD 20%), respectively. Sleep tracking had much lower compliance rates, with an average of 47% (SD 37%) compliance. One would expect the passively collected measures to have the highest compliance, which was true in our study, except for assessment of sleep. Sleep was measured via the free Sleep++ app that was downloaded onto the participants' Apple Watches. The largest number of complaints were about the Sleep++ app, which participants found to be difficult to operate, would not track sleep accurately, or would not initiate sleep tracking despite starting the tracking process. Our feedback questionnaire also asked explicitly about overnight comfort of wearing the watch ("Was it comfortable to wear the Apple Watch while sleeping?") and received feedback from 8 participants that it was not because of the bulk of the watch, alerts or lighting up at night, or general discomfort while wearing jewelry or watches overnight. We believe that this contributed to the reduction in compliance for sleep tracking.

As part of assessing the clinical utility of the data collected, the predictive value of these data were examined. The preliminary data on prediction are interesting and possibly of clinical value. These small-scale associations show that EMA ratings of mood and anxiety in between clinical appointments are correlated with validated scales for mood and anxiety, which could be highly useful in assessing the trajectory of a patient's condition in the periods between clinical contacts. This could be especially useful for postpartum patients who have many barriers to treatment and remission for PPD [1,23]. This type of EMA tracking could be used to signal both patients and clinicians that mood or anxiety is worsening, prompting initiation or changes in the treatment plan. With the decrease in clinical care appointments due to COVID-19, this type of monitoring could be used to determine when appointments are needed and be an additional safety check for postpartum women in between scheduled check-ins. These associations will need further investigation, as significance levels may be driven by the small sample size; however, the trends are encouraging.

Barriers to Adoption and Continued Use

Two of the major issues related to acceptance were the size of the Apple Watch and use of the Sleep++ app. The Apple Watches used for the study were the larger of the Series 2, with a 42-mm face (vs 38 mm). This larger face watch size likely

could have contributed to the discomfort experienced by some participants. In addition, participants reported many technical issues with the Sleep++ app, which was chosen because there was no extra cost to build sleep monitoring functionality into *PPD ACT* and the data were easily integrated using Apple CareKit. At the time the module was developed, integrated sleep tracking was not available with the Apple Watch or Health Kit. Sleep tracking has advanced rapidly since the completion of the pilot study, and integrated sleep tracking is available on many wearable devices, including the Apple Watch. Most sleep trackers now intuitively track the wearer's sleep, as long as the device is worn to bed. The other noted acceptance issue is regarding the process of tracking itself and the possibility that for some, tracking may exacerbate existing anxieties. To mitigate these concerns, clinicians should regularly check in with patients to determine whether their symptoms are directly affected, either positively or negatively, by the use of this tracking modality.

Limitations and Future Directions

Despite promising initial data on the feasibility of using EMA to enhance clinical insight among women experiencing PPD, there are limitations to this pilot study. First, all but 1 participant who enrolled were already in treatment for PPD, so no insight can be gained as to the efficacy of this technology in helping initiate PPD treatment. Furthermore, as the study clinician was not the participants' usual care provider, the study clinician was not able to draw conclusions about the impact of EMA across the course of treatment. Although the use of EMA proved feasible over the course of the 6-week study, we do not know the long-term effect or acceptability of this technology beyond the study period. A larger study of this technology, incorporating the lessons learned from this pilot study, would provide additional information about the long-term use of tracking among women in treatment for PPD and the predictive value of daily mood and anxiety assessments. This information may benefit women in areas of the country where dedicated perinatal psychiatry programs do not exist and assist in determining appropriate treatment modalities and intervals of treatment.

Despite these limitations, valuable information was gained from this pilot study, including a clear demonstration of the acceptability, feasibility, and clinical utility of EMA tracking among women experiencing PPD. These lessons are even more valuable in the context of the pandemic.

Conclusions

EMA has not been studied to date as a method for the enhancement of clinical care in postpartum women. This study found EMA tracking to be acceptable among participants and was endorsed as clinically useful by the study psychiatrist. Given the barriers to care faced by many women with PPD, this largely home-based technology could help both women and their providers better understand the trajectory of their symptoms and identify areas where improvements could be made in the management of their mental health needs. This need is all the greater with the crisis COVID-19 has caused in perinatal mental health.

Multimedia Appendix 1
Apple Watch Module screens.
[PDF File (Adobe PDF File), 2506 KB - formative_v6i3e28081_app1.pdf]

Multimedia Appendix 2
Apple Watch study feedback surveys for clinician and participants.
[PDF File (Adobe PDF File), 1286 KB - formative_v6i3e28081_app2.pdf]

References

1. Cox EQ, Sowa NA, Meltzer-Brody SE, Gaynes BN. The perinatal depression treatment cascade. J Clin Psychiatry 2016 Sep 28;77(09):1189-1200. [doi: 10.4088/jcp.15r10174]
2. Stewart DE, Vigod S. Postpartum depression. N Engl J Med 2016 Dec;375(22):2177-2186. [doi: 10.1056/nejmcp1607649]
3. Gaynes B, Gavin N, Meltzer-Brody S, Lohr K, Swinson T, Gartlehner G, et al. Perinatal depression: prevalence, screening accuracy, and screening outcomes. Evid Rep Technol Assess (Summ) 2005 Feb(119):1-8. [doi: 10.1037/e439372005-001] [Medline: 15760246]
4. Gavin N, Gaynes B, Lohr K, Meltzer-Brody S, Gartlehner G, Swinson T. Perinatal depression. Obstetrics Gynecol 2005;106(5, Part 1):1071-1083. [doi: 10.1097/01.aog.0000183597.31630.db]

XSL•FO

RenderX

5. Lindahl V, Pearson JL, Colpe L. Prevalence of suicidality during pregnancy and the postpartum. Arch Womens Ment Health 2005 Jun 11;8(2):77-87. [doi: 10.1007/s00737-005-0080-1] [Medline: 15883651]

6. Dagher R, McGovern P, Dowd B, Gjerdingen D. Postpartum depression and health services expenditures among employed women. J Occup Environ Med 2012 Feb;54(2):210-215. [doi: 10.1097/JOM.0b013e31823fdf85] [Medline: 22267187]

7. Dennis C, McQueen K. The relationship between infant-feeding outcomes and postpartum depression: a qualitative systematic review. Pediatrics 2009 Apr;123(4):e736-e751. [doi: 10.1542/peds.2008-1629] [Medline: 19336362]

8. Earls MF, Yogman MW, Mattson G, Rafferty J, COMMITTEE ON PSYCHOSOCIAL ASPECTS OF CHILD AND FAMILY HEALTH. Incorporating recognition and management of perinatal depression into pediatric practice. Pediatrics 2019 Jan;143(1):e20183260. [doi: 10.1542/peds.2018-3259] [Medline: 30559120]

9. Netsi E, Pearson RM, Murray L, Cooper P, Craske MG, Stein A. Association of persistent and severe postnatal depression with child outcomes. JAMA Psychiatry 2018 Mar 01;75(3):247-253 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.4363] [Medline: 29387878]

10. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008 Apr 01;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]

11. Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. Psychol Assess 2009 Dec;21(4):463-475. [doi: 10.1037/a0017075] [Medline: 19947781]

12. Asselbergs J, Ruwaard J, Ejdys M, Schrader N, Sijbrandij M, Riper H. Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study. J Med Internet Res 2016 Mar 29;18(3):e72 [FREE Full text] [doi: 10.2196/jmir.5505] [Medline: 27025287]

13. Kim H, Lee S, Lee S, Hong S, Kang H, Kim N. Depression prediction by using ecological momentary assessment, Actiwatch data, and machine learning: observational study on older adults living alone. JMIR Mhealth Uhealth 2019 Oct 16;7(10):e14149 [FREE Full text] [doi: 10.2196/14149] [Medline: 31621642]

14. Hung S, Li M, Chen Y, Chiang J, Chen Y, Hung G. Smartphone-based ecological momentary assessment for Chinese patients with depression: an exploratory study in Taiwan. Asian J Psychiatr 2016 Oct;23:131-136. [doi: 10.1016/j.ajp.2016.08.003] [Medline: 27969071]

15. Kim J, Uddin ZA, Lee Y, Nasri F, Gill H, Subramanieapillai M, et al. A systematic review of the validity of screening depression through Facebook, Twitter, Instagram, and Snapchat. J Affect Disord 2021 May 01;286:360-369. [doi: 10.1016/j.jad.2020.08.091] [Medline: 33691948]

16. Heller AS, Fox AS, Davidson RJ. Parsing affective dynamics to identify risk for mood and anxiety disorders. Emotion 2019 Mar;19(2):283-291 [FREE Full text] [doi: 10.1037/emo0000440] [Medline: 29863379]

17. Sanjuan PM, Pearson MR, Poremba C, Amaro HD, Leeman L. An ecological momentary assessment study examining posttraumatic stress disorder symptoms, prenatal bonding, and substance use among pregnant women. Drug Alcohol Depend 2019 Feb 01;195:33-39 [FREE Full text] [doi: 10.1016/j.drugalcdep.2018.11.019] [Medline: 30572290]

18. Mendez D, Sanders S, Karimi H, Gharani P, Rathbun S, Gary-Webb T, et al. Understanding pregnancy and postpartum health using ecological momentary assessment and mobile technology: protocol for the postpartum mothers mobile study. JMIR Res Protoc 2019 Jun 26;8(6):e13569. [doi: 10.2196/13569]

19. Giesbrecht GF, Letourneau N, Campbell T, Kaplan BJ, APrON Study Team. Affective experience in ecologically relevant contexts is dynamic and not progressively attenuated during pregnancy. Arch Womens Ment Health 2012 Dec;15(6):481-485. [doi: 10.1007/s00737-012-0300-4] [Medline: 22926981]

20. Osborne LM, Kimmel MC, Surkan PJ. The crisis of perinatal mental health in the age of covid-19. Matern Child Health J 2021 Mar 04;25(3):349-352 [FREE Full text] [doi: 10.1007/s10995-020-03114-y] [Medline: 33543374]

21. Meltzer-Brody S, Kimmel M. The promise of telepsychiatry to reduce maternal mortality by increasing access to maternal mental health and addiction services. Obstet Gynecol 2020 Oct;136(4):643-644. [doi: 10.1097/AOG.0000000000004116] [Medline: 32925618]

22. Guintivano J, Krohn H, Lewis C, Byrne EM, Henders AK, Ploner A, et al. PPD ACT: an app-based genetic study of postpartum depression. Transl Psychiatry 2018 Nov 29;8(1):260 [FREE Full text] [doi: 10.1038/s41398-018-0305-5] [Medline: 30498212]

23. Dennis C, Chung-Lee L. Postpartum depression help-seeking barriers and maternal treatment preferences: a qualitative systematic review. Birth 2006 Dec;33(4):323-331. [doi: 10.1111/j.1523-536x.2006.00130.x]

24. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 1987 Jun;150:782-786. [doi: 10.1192/bjp.150.6.782] [Medline: 3651732]

25. Meltzer-Brody S, Boschloo L, Jones I, Sullivan PF, Penninx BW. The EPDS-Lifetime: assessment of lifetime prevalence and risk factors for perinatal depression in a large cohort of depressed women. Arch Womens Ment Health 2013 Dec 1;16(6):465-473 [FREE Full text] [doi: 10.1007/s00737-013-0372-9] [Medline: 23904137]

26. Santos I, Tavares B, Munhoz T, Manzolli P, de Ávila GB, Jannke E, et al. Patient health questionnaire-9 versus Edinburgh postnatal depression scale in screening for major depressive episodes: a cross-sectional population-based study. BMC Res Notes 2016 Sep 27;9(1):453 [FREE Full text] [doi: 10.1186/s13104-016-2259-0] [Medline: 27677844]

27. Barrios YV, Gelaye B, Zhong Q, Nicolaidis C, Rondon MB, Garcia PJ, et al. Association of childhood physical and sexual abuse with intimate partner violence, poor general health and depressive symptoms among pregnant women. PLoS One 2015 Jan 30;10(1):e0116609 [FREE Full text] [doi: 10.1371/journal.pone.0116609] [Medline: 25635902]

28. Yawn BP, Pace W, Wollan PC, Bertram S, Kurland M, Graham D, et al. Concordance of Edinburgh Postnatal Depression Scale (EPDS) and Patient Health Questionnaire (PHQ-9) to assess increased risk of depression among postpartum women. J Am Board Fam Med 2009 Sep 04;22(5):483-491 [FREE Full text] [doi: 10.3122/jabfm.2009.05.080155] [Medline: 19734393]

29. Levis B, Benedetti A, Thombs B, DEPRESsion Screening Data (DEPRESSD) Collaboration. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ 2019 Apr 09;365:l1476 [FREE Full text] [doi: 10.1136/bmj.l1476] [Medline: 30967483]

30. Maust D, Cristancho M, Gray L, Rushing S, Tjoa C, Thase M. Psychiatric rating scales. In: Handbook of Clinical Neurology. Amsterdam, Netherlands: Elsevier; 2012.

31. Kroenke K, Spitzer R, Williams J, Löwe B. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. Gen Hosp Psychiatry 2010;32(4):345-359. [doi: 10.1016/j.genhosppsych.2010.03.006] [Medline: 20633738]

32. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]

33. Cox JL, Chapman G, Murray D, Jones P. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in non-postnatal women. J Affect Disord 1996 Jul 29;39(3):185-189. [doi: 10.1016/0165-0327(96)00008-0] [Medline: 8856422]

34. Barkin JL, Wisner KL, Bromberger JT, Beach SR, Terry MA, Wisniewski SR. Development of the Barkin index of maternal functioning. J Womens Health (Larchmt) 2010 Dec;19(12):2239-2246 [FREE Full text] [doi: 10.1089/jwh.2009.1893] [Medline: 21054183]

35. Barkin JL, Wisner KL, Wisniewski SR. The psychometric properties of the Barkin Index of Maternal Functioning. J Obstet Gynecol Neonatal Nurs 2014;43(6):792-802 [FREE Full text] [doi: 10.1111/1552-6909.12505] [Medline: 25315473]

36. Braun A, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

37. Applied Thematic Analysis. Thousand Oaks, California: SAGE Publications; 2012.

38. Boyd RC, Le HN, Somberg R. Review of screening instruments for postpartum depression. Arch Womens Ment Health 2005 Sep;8(3):141-153. [doi: 10.1007/s00737-005-0096-6] [Medline: 16133785]

39. Hewitt C, Gilbody S. Is it clinically and cost effective to screen for postnatal depression: a systematic review of controlled clinical trials and economic evidence. BJOG 2009 Jul;116(8):1019-1027. [doi: 10.1111/j.1471-0528.2009.02148.x] [Medline: 19438499]

40. Johnson EI, Grondin O, Barrault M, Faytout M, Helbig S, Husky M, et al. Computerized ambulatory monitoring in psychiatry: a multi-site collaborative study of acceptability, compliance, and reactivity. Int J Methods Psychiatr Res 2009 Mar;18(1):48-57 [FREE Full text] [doi: 10.1002/mpr.276] [Medline: 19195050]

41. Depp CA, Kim DH, de Dios LV, Wang V, Ceglowski J. A pilot study of mood ratings captured by mobile phone versus paper-and-pencil mood charts in bipolar disorder. J Dual Diagn 2012 Jan 01;8(4):326-332 [FREE Full text] [doi: 10.1080/15504263.2012.723318] [Medline: 23646035]

42. Juengst SB, Graham KM, Pulantara IW, McCue M, Whyte EM, Dicianno BE, et al. Pilot feasibility of an mHealth system for conducting ecological momentary assessment of mood-related symptoms following traumatic brain injury. Brain Inj 2015;29(11):1351-1361. [doi: 10.3109/02699052.2015.1045031] [Medline: 26287756]

Abbreviations

BIMF: Barkin Index of Maternal Functioning
EMA: ecological momentary assessment
EPDS: Edinburgh Postnatal Depression Scale
GAD-7: Generalized Anxiety Disorder 7-item scale
PHQ-9: Patient Health Questionnaire-9
PPD: postpartum depression
RMSE: root-mean-square error

XSL·FO

RenderX

Original Paper

Pilot Results of a Digital Hypertension Self-management Program Among Adults With Excess Body Weight: Single-Arm Nonrandomized Trial

Folasade Wilson-Anumudu[1], MPH; Ryan Quan[1], MPH; Christian Cerrada[2], PhD; Jessie Juusola[2], PhD; Cynthia Castro Sweet[1], PhD; Carolyn Bradner Jasik[1], MD; Michael Turken[1], MD, MPH

[1]Omada Health, Inc, San Francisco, CA, United States
[2]Evidation Health, Inc, San Mateo, CA, United States

Corresponding Author:
Folasade Wilson-Anumudu, MPH
Omada Health, Inc
500 Sansome Street, Suite 200
San Francisco, CA, 94111
United States
Phone: 1 6502696532
Email: folasade.anumudu@omadahealth.com

Abstract

Background: Home-measured blood pressure (HMBP) in combination with comprehensive medication support and lifestyle change are the mainstays of evidence-based hypertension (HTN) management. To date, the precise components needed for effective HTN self-management programs have yet to be defined, and access to multicomponent targeted support for HTN management that include telemonitoring remain inaccessible and costly.

Objective: The aim of this pilot study was to evaluate the impact of a digital HTN self-management program on blood pressure (BP) control among adults with excess body weight.

Methods: A single-arm, nonrandomized trial was performed to evaluate a digital HTN self-management program that combines comprehensive lifestyle counseling with HTN education, guided HMBP, support for taking medications, and led by either a registered nurse or certified diabetes care and education specialist. A sample of 151 participants were recruited using a web-based research platform (Achievement Studies, Evidation Health Inc). The primary outcome was change in systolic BP from baseline to 3 months, and secondary outcomes included change in diastolic BP and medication adherence.

Results: Participants' mean age was 44.0 (SD 9.3) years and mean BP was 139/85 mm Hg. At follow-up, systolic and diastolic BP decreased by 7 mm Hg ($P<.001$, 95% CI –9.3 to –4.7) and 4.7 mm Hg ($P<.001$, 95% CI –6.3 to –3.2), respectively. Participants who started with baseline BP at goal remained at goal. For participants with stage 1 HTN, systolic and diastolic BP decreased by 3.6 mm Hg ($P=.09$, 95% CI –7.8 to 0.6) and 2.5 mm Hg ($P=.03$, 95% CI –4.9 to –0.3). Systolic and diastolic BP decreased by 10.3 mm Hg ($P<.001$, 95% CI –13.4 to –7.1) and 6.5 mm Hg ($P<.001$, 95% CI –8.6 to –4.4), respectively, for participants with stage 2 HTN. Medication adherence significantly improved ($P=.02$).

Conclusions: This pilot study provides initial evidence that a digital HTN self-management program improves BP and medication adherence.

Introduction

Background

Hypertension (HTN) impacts an estimated 116.4 million or 46% of adults in the United States [1]. Annual health care costs associated with HTN are approximately US $131 billion or 3% of the US $3 trillion national health care expenditure [2]. The proportion of adults with controlled blood pressure (BP), has declined in recent years, with <44% of adults achieving this goal [3]. Poor BP control continues to be a persistent public

health challenge because of the high prevalence of obesity, insufficient access to lifestyle counseling, difficulty adopting lifestyle change, medication adherence, therapeutic inertia owing to clinical uncertainty about BP data accuracy, and lack of provider access to adjust medications when needed [4-7].

Home-measured blood pressure (HMBP), when combined with more intensive human-led interventions, such as lifestyle counseling or medication titration, has shown promise in improving BP control [8]. In particular, HMBP data may be superior to in-office measurement to inform treatment decisions. The most effective interventions seem to be those that combine HMBP with medication counseling to address medication-taking beliefs and barriers and facilitate rapid medication titration [9,10]. These programs are typically staffed by pharmacists and embedded within clinics, making them difficult to scale efficiently. Clinics struggle with maximizing pharmacist time and dispensing or triaging devices.

Mobile health (mHealth) interventions can facilitate scalable, accessible, and effective management of HTN. HMBP is a core component with additional features such as automated medication reminders [11] or artificial intelligence (AI) software to provide real-time feedback on monitored BP values [12]. The bulk of innovation in this area has focused on fully automated solutions, but these interventions have had marginal clinical efficacy and have not been shown to be more clinically effective than HMBP alone. Few programs combine HMBP with access to what is known to work—human-led support that interprets BP data, supports lifestyle change, and titrates medication.

Self-management is a daily process where individuals actively engage in the management of a chronic illness, and structured programs such as Diabetes Self-Management Education and Support have demonstrated broad clinical efficacy [13]. Similarly, for HTN, supported self-management interventions, which are complex and employ a broad range of support strategies, have been found to improve BP control [14]. While robust evidence is still lacking, early evidence suggests that mHealth interventions with more comprehensive features are likely to be more effective [15,16]. The self-management approach places focus on the individual to play a critical role in their care in collaboration with health care providers [17].

Objective

The goal of this pilot study was to evaluate the impact on BP control of a registered nurse (RN) or certified diabetes care and education specialist (CDCES)-led mHealth HTN self-management program that combines comprehensive lifestyle counseling (for dietary changes, weight loss, physical activity, and stress management) with HTN education, guided HMBP, support for taking medications, and social support, among individuals with uncontrolled HTN.

Methods

Participants

Members of a web-based health community (Achievement, Evidation Health Inc) were invited to participate in this study. Achievement is a web- and mobile-based community in the United States where members can connect their activity trackers

and fitness and health apps to the platform and, by logging activities, accumulate points that are redeemable for monetary rewards. Additionally, members self-report on various health conditions and are invited to participate in remote research opportunities as relevant studies come available. In this study, recruitment was targeted to members who had self-reported an HTN diagnosis. Invited members were linked to a web-based research study platform (Achievement Studies, Evidation Health Inc) where study eligibility was assessed using automated screener questions. Eligibility criteria for the pilot study included the following: being a US resident, being at least 18 years of age, self-reported HTN diagnosis, self-reported recent systolic BP (SBP) measurement of ≥130 mm Hg or diastolic BP (DBP) of ≥80 mm Hg within 1 month prior to screening, a BMI of ≥25 kg/m^2 (≥23 kg/m^2 if they self-identified as Asian), and access to a computer or smartphone to participate in the virtual HTN program.

Procedures

If eligible after completing the web-based screener, potential participants were asked to sign an electronic informed consent form and complete a web-based baseline survey, consisting of questions about their demographics, health and HTN history, HTN medication usage, and patient-reported outcomes (as described in *Measurements*). After completion of the baseline survey, potential participants were instructed to set up a HTN program account. After completion of both the baseline survey and program account setup, participants were considered to have been enrolled in the study. In the first week of the program, participants were instructed to take BP readings and were asked to take multiple readings throughout the week. The average across all readings in the first 30 days was used as the baseline measurement for the pilot study.

Throughout the study period, participants were encouraged to engage with the supported mHealth HTN self-management program at the time and frequency of their choosing. After 3 months of study enrollment, participants were asked to complete a final web-based survey and submit BP measurements throughout the last week of the program. Study participation was completely remotely, and participants received a small electronic gift card compensation for completion of all data collection.

Ethical Considerations

This study was exempted from ethics approval by the Western International Review Board (WIRB Work Order #1-1249356-1). All participants in this study provided informed consent.

Measurements

All study measures were collected at baseline and at 3 months. BP was collected using a cellular-connected BP monitor (BodyTrace Inc) that was provided to every participant as a core component of the program. Participants were provided with instructions on how to take accurate resting BP readings [18]. Participants were asked to submit a BP measurement at baseline and at 3 months. Baseline BP was defined as the average of at least 3 BP readings taken across 2 days in a 30-day measurement window closest to the baseline time point. Similarly, final BP was defined as the average of at least 3 BP readings taken across

2 days in a 30-day measurement window closest to the 3-month time point. Throughout the study, participants were encouraged to submit additional BP measurements in accordance with the schedule recommended by their health care professional. Program participants were also provided with a cellular-connected weight scale (BodyTrace Inc) and were asked to weigh in daily throughout the program. The median weight collected during weeks 1 and 12 of the program was used as weight outcomes for the pilot study.

The following patient-reported outcomes were assessed through a web-based survey administered at baseline and during week 12 of the program: the Consumer Health Activation Index (CHAI), a 10-item scale measuring health-related activation and engagement [19], with scores between 0 and 79 reflect low activation, scores between 80 and 94 reflecting moderate activation, and 95 and 100 reflecting high activation [19]; the Self-Efficacy for Managing Chronic Disease scale, 6 items on a scale of 1 (not at all confident) to 10 (totally confident) measuring one's confidence in doing certain activities that are common across many chronic conditions, with higher scores indicating higher self-efficacy [20]; HTN medication usage (ie, a self-report of medications, dosage, and timing of administration); and the Simplified Medication Adherence Questionnaire (SMAQ), a 6-item scale that categorizes respondents as adherent or nonadherent on the basis of recent patterns of medication-taking behaviors [21].

Intervention

The Omada for Hypertension Program is a supported digital HTN self-management program that is commercially available through business-to-business relationships with organizations that pay to make the service available to their members (typically employees or health plan members). The program can be accessed through mobile devices (ie, smartphones or tablet devices) or PCs. The program offers a HTN education curriculum along with comprehensive lifestyle self-management support, including support for weight loss, dietary changes (aligning with the Dietary Approaches to Stop Hypertension and Mediterranean Diets), increased physical activity, support for medication adherence, social support, and a cellularly connected BP monitor and body weight scale. Participants are assigned to a HTN specialist coach (either a RN or CDCES), who supports their progress longitudinally through the program, addressing specific questions, providing feedback on HMBP data, supporting medication-taking, and helping prepare participants for primary care physician and specialist visits. Data from the BP monitor and weight scale are provided back to the participant in their program account; the data are also used by the HTN specialist coach to provide counseling on dietary changes, changes in physical activity, and to encourage communication with their health care providers when timely medication adjustments may be needed. Participants are also placed in a virtual peer group and can communicate with other program users through a secure group discussion board.

Statistical Analysis

The study was powered to detect a clinically meaningful 4 mm Hg reduction in the primary outcome of resting SBP. With an estimated SD of 12 mm Hg and power set to 90%, the minimum sample size needed was 113. To allow for a potential 10% loss to follow-up of study participants at 3 months, and an estimate of 30% of participants' resting SBP being <130 mm Hg at baseline, a total of 150 participants were planned for study enrollment.

Descriptive statistics (eg, means and frequencies) were performed to describe the demographic and baseline clinical characteristics of the enrolled study population. Correlations (Pearson correlation for continuous variables and Spearman correlation for categorical variables) were run to detect potential baseline confounders (age, gender, and BMI) of resting SBP and DBP; no significant correlations were found. Two-tailed paired t tests were used to assess significant changes in BP measurements, weight, and patient-reported outcomes from baseline to follow-up. The definitions of stage 1 and stage 2 HTN differ in accordance with BP measurement methodology, with the HMBP threshold for stage 2 HTN (135/85 mm Hg) being lower than the office-based threshold (140/90 mm Hg) [22]. While the HMBP threshold is the most pertinent for this study given the BP measurement methodology, the majority of trials have traditionally used office-based thresholds. Hence, for BP post hoc analyses, participants were stratified by HTN stage using both home-based and in-office cutoffs for stage 1 and stage 2 HTN, and paired t tests were performed within these groups to examine changes in BP on the basis of the baseline BP stage.

The McNemar nonparametric test was used to examine the change in the proportion of the study population that was adherent to their medication regimen from baseline to follow-up and 2-tailed paired t tests were used to assess change in the number of HTN medications used by participants.

Outcomes were analyzed using both intention-to-treat analysis (with baseline carry forward of missing data) and complete case analysis (among those with complete baseline and follow-up data). The outcomes were similar in magnitude and statistical significance using both analytic methods, and thus we present the results on the sample of study participants with complete data from both timepoints.

Results

Study Recruitment

The final enrolled sample comprised 153 participants. Three participants were withdrawn from the study: one developed a medical condition that precluded participation and 2 requested to voluntarily withdraw from the study. At follow-up, 80% (n=121) of participants had a complete follow-up BP measurement, and 81% completed the web-based questionnaire; 29 (19%) participants were lost to follow-up. Figure 1 outlines the flow of participants through each stage of the study. The analyzed sample (n=121) and those lost to follow-up (n=30) were compared for differences in baseline characteristics, which might impact the BP outcomes. There were no significant differences in baseline SBP, weight, medication adherence, gender, age, or race and ethnicity. However, baseline DBP was slightly higher in the group lost to follow-up than in those who stayed in the study (88 vs 84 mm Hg).

XSL·FO

RenderX

Figure 1. Flow diagram of study participants through each stage of the study.
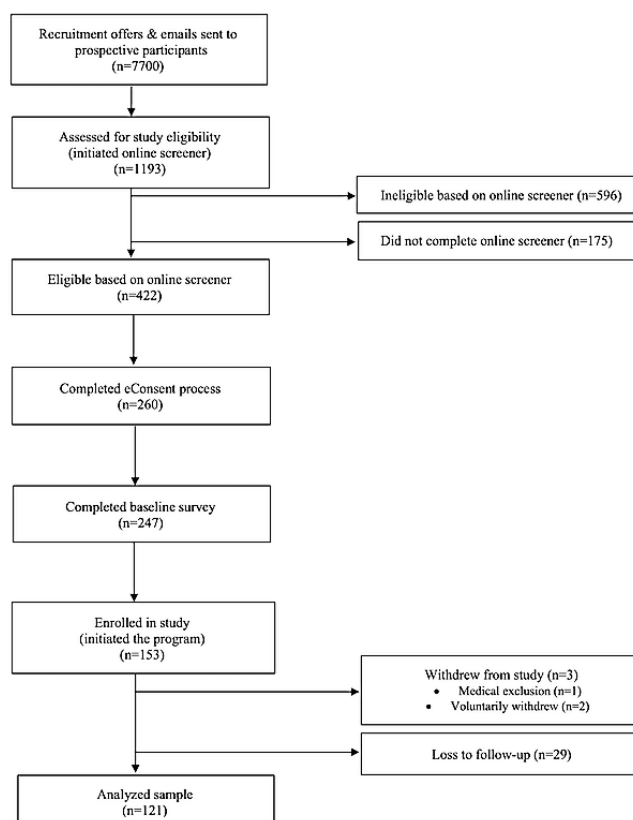
Participant Characteristics at Baseline

Table 1 shows the baseline characteristics of the study participants, including missing data for participants who did not complete various metrics at baseline. The sample was 56% female, 72% White, and 11% of participants self-identifying as African American. The baseline mean SBP for the sample was 139 mm Hg and mean DBP 85 was mm Hg. In total, 76% of participants reported taking medication for HTN medication at the start of the program and 27% self-reported being adherent to their medication regimen (SMAQ).

The average weight of the sample at program start was 224 lbs. The mean CHAI health activation score at baseline was 71.9 (SD 14.2). The mean self-efficacy score at baseline was 44.3 (SD 11.7) out of a possible 60.

Table 1. Baseline participant characteristics (N=151).

Baseline characteristics[a]	Values
Age (years), mean (SD)	44.0 (9.3)
Females, n (%)	85 (56.3)
White participants, n (%)	109 (72.2)
Black or African American participants, n (%)	17 (11.3)
Systolic blood pressure (mm Hg; n=148), mean (SD)	138.5 (13.1)
Diastolic blood pressure (mm Hg; n=148), mean (SD)	84.6 (8.9)
Weight (lbs; n=149), mean (SD)	223.7 (52.0)
Patient activation (n=144), mean (SD)	71.9 (14.2)
Self-efficacy (n=144), mean (SD)	44.3 (11.7)
Taking hypertension medication, n (%)	114 (75.5)
Adherent to current medications, n (%)	41 (27.2)

[a]The table includes missing data from participants who did not complete baseline data collection.

Program Engagement

Participants used their BP cuff an average of 7.2 times per week across the 12 weeks of the program. Participants weighed in an average of 4.7 times per week, interacted with their coaches an average of 1.2 times per week, completed an average of 0.75 lessons per week, interacted with their peer groups an average of 0.6 times per week, tracked their physical activity 5.4 times per week, and tracked meals an average of 9.0 times per week.

BP Outcomes

In Table 2, the baseline and follow-up changes in mean BP are listed for the analyzed sample (n=121). Baseline SBP significantly decreased by an average of 7 mm Hg (t_{120}=–6.0; P<.001; 95% CI –9.3 to –4.7). DBP significantly declined by 4.7 mm Hg (t_{120}=–6.1; P<.001; 95% CI –6.3 to –3.2). Weight also decreased by an average of –2.4 lbs (t_{119}=–3.5; P<.001; 95% CI –3.7 to –1.0), with 14% of the sample losing 5% of their initial body weight. Table 3 presents changes in BP by baseline BP clinical category based on home-based versus

in-office thresholds for stage 2 HTN [22]. All study participants who started with BP at goal remained at goal (26/121, 21.5%). For participants with stage 1 HTN at baseline (SBP 130-134 mm Hg, DBP 80-84 mm Hg), there was a trend toward significant reduction in SBP by 3.6 mm Hg (t_{18}=–1.8; P=.09; 95% CI –7.8 to 0.6), and DBP significantly decreased by 2.6 mm Hg (t_{18}=–2.4; P=.03; 95% CI –4.9 to –0.3). BP reduction was greater for those with stage 2 HTN at baseline (SBP≥135 mm Hg or DBP≥85 mm Hg), with a 10.3 mm Hg reduction in SBP (t_{75}=–6.5; P<.001; 95% CI –13.4 to –7.1) and a 6.5 mm Hg reduction in DBP (t_{75}=–6.1; P<.001; 95% CI –8.6 to –4.4). Using in-office range thresholds, those with stage 1 HTN (SBP 130-139 mm Hg, DBP 80-89 mm Hg) had a significantly decreased DBP by 2.9 mm Hg (t_{38}=–3.1; P<.004; 95% CI –4.8 to –0.98). For stage 2 HTN defined using office ranges (SBP≥140 mm Hg or DBP≥90 mm Hg), SBP reduction was 13.4 mm Hg (t_{55}=–7.3; P<.001; 95% CI –17.1 to –9.7) and a 7.7 mm Hg reduction in DBP (t_{55}=–5.8; P<.001; 95% CI –10.4 to –5.1).

Table 2. Baseline to follow-up changes in outcomes in participants with complete data.

Parameters	Participants, n	Baseline value	Follow-up value	Difference	t test (df)	95% CI	P value
Systolic blood pressure (mm Hg)	121	137.7	130.7	–7.0	–6.0 (120)	–9.3 to –4.7	<.001
Diastolic blood pressure (mm Hg)	121	83.9	79.1	–4.7	–6.1 (120)	–6.3 to –3.2	<.001
Weight (lbs)	120	221.8	219.4	–2.4	–3.5 (119)	–3.7 to –1.0	.001
Patient activation	115	70.9	74.0	3.1	2.4 (114)	0.6 to 5.6	.02
Self-efficacy	115	43.9	44.4	0.4	0.5 (114)	–1.5 to 2.3	.65

Table 3. Baseline to follow-up changes in blood pressure by the baseline blood pressure category from among participants with complete data.

Parameters	Participants, n	Baseline value, mm Hg	Follow-up value, mm Hg	Difference, mm Hg	t test (df)	95% CI	P value
Home range							
Systolic blood pressure							
Blood pressure at goal	26	122.0	122.0	0.0	0.0 (25)	–3.4 to 3.4	1.00
Stage 1 hypertension	19	130.0	126.0	–3.6	–1.8 (18)	–7.8 to 0.6	.09
Stage 2 hypertension	76	145.0	135.0	–10.3	–6.5 (75)	–13.4 to –7.1	<.001
Diastolic blood pressure							
Blood pressure at goal	26	74.0	73.0	–1.1	–0.9 (25)	–3.7 to 1.5	.40
Stage 1 hypertension	19	79.0	76.4	–2.6	–2.4 (18)	–4.9 to –0.3	.03
Stage 2 hypertension	76	88.4	81.9	–6.5	–6.1 (75)	–8.6 to –4.4	<.001
Office range							
Systolic blood pressure							
Blood pressure at goal	26	122.0	122.0	0.0	0.0 (25)	–3.4 to 3.4	1.00
Stage 1 hypertension	39	133.0	130.0	–2.5	–1.7 (38)	–5.4 to 0.5	.10
Stage 2 hypertension	56	148.0	135.0	–13.4	–7.3 (55)	–17.1 to –9.7	<.001
Diastolic blood pressure							
Blood pressure at goal	26	74.0	73.0	–1.1	–0.9 (25)	–3.7 to 1.5	.40
Stage 1 hypertension	39	81.4	78.5	–2.9	–3.1 (38)	–4.8 to –0.98	.004
Stage 2 hypertension	56	90.1	82.4	–7.7	–5.8 (55)	–10.4 to –5.1	<.001

Medication Outcomes

Table 4 presents changes in medication status and medication adherence from baseline to follow-up. Medication adherence, as measured by the SMAQ, improved for the total sample by 14.4% from 36.7% to 51.1% (McNemar $\chi^2_{1,90}$=5.8, P=.02).

When we conducted post hoc analyses by HTN stage, there were no significant changes in the number of medications among those who began the program at goal or with stage 1 HTN. However, the number of HTN medications increased from 1.0 to 1.2 (t_{76}=2.2; P=.03; 95% CI 0.01-0.30) among those with baseline stage 2 HTN based on the home-based thresholds.

Table 4. Baseline to follow-up changes in medication status.

Parameters	Participants, n	Baseline value	Follow-up value	Difference	t test (df)	McNemar χ^2 (df)	95% CI	P value
Total sample								
Average number of hypertension medications	117	1.1	1.2	0.1	1.1 (116)	N/A[a]	–0.05 to 0.17	.28
Adherent to medications (Simplified Medication Adherence Questionnaire), %	90[b]	36.7	51.1	14.4	N/A	5.8 (90)	N/A	.02
Stage 2 subsample								
Average number of hypertension medications	77	1.0	1.2	0.2	2.2 (76)	N/A	0.01 to 0.30	.03
Adherent to medications (Simplified Medication Adherence Questionnaire), %	54[b]	35.2	46.3	11.1	N/A	2.0 (54)	N/A	.16

[a]N/A: not applicable.

[b]Among participants taking ≥1 hypertension medications at baseline.

Patient-Reported Outcomes

Patient activation on the CHAI significantly increased by an average of 3.1 points (t_{114}=2.4; P=.02; 95% CI 0.6-5.6) but self-efficacy did not improve significantly (P=.65).

Discussion

Principal Findings

The results of this pilot study provide initial evidence that a comprehensive, human-led digital HTN self-management program that includes lifestyle support, medication adherence, and guided HMBP is associated with improved BP control, weight, and medication adherence in a sample of individuals with uncontrolled HTN. Furthermore, those who started the program with stage 2 HTN achieved the greatest improvement in BP control with an average change of 10.3 mm Hg and 6.5 mm Hg in SBP and DBP, respectively. Those who started the program with their BP at goal remained at goal at the end of the study. This pilot study was successful in detecting significant weight loss and improvement in medication adherence. Program engagement was strong, as shown by the high frequency of use across the various features of the digital platform.

The findings from this pilot study are consistent with those of prior studies on digital HTN programs that showed improvements in BP control, weight, and medication adherence [11,16,23,24]. While most participants saw significant improvements in their BP, those with stage 2 HTN at baseline saw the greatest improvement, with a magnitude of decrease similar to those seen with interventions that are led by pharmacists or nurse practitioners [9]. Additionally, the

magnitude of BP reduction observed in this program is comparable to that of previous studies [25,26].

There were no significant changes seen in self-efficacy; this may be in part owing to participants reporting relatively high self-efficacy at baseline. There were significant improvements in patient activation, which is arguably more comprehensive and encompassing than self-efficacy for managing HTN [27].

Limitations

The single-arm, nonrandomized design of this pilot study harbors the challenge of unknown causal inference; thus, future studies with a comparison or control group are needed to confirm the results. Second, all measurements were collected at home using cellularly connected devices. While there is a risk of small intradevice measurement variability, participants did use the same device for their measurements across the study, so any measurement error is likely to be systematic within individuals. Third, participants self-selected from the web-based health community into the research opportunity; therefore, it is possible that the study population recruited for the pilot study is not fully representative or generalizable of the population of individuals living with HTN in the United States. The pilot study was conducted from February to July 2020 during the first wave of COVID-19 exposure. This may have influenced participants' ability to fully engage in the program in the midst of unprecedented stressors and disruptions in health care. However, the observation of significant clinical benefits during this time is an encouraging and suggests that the results might have been stronger if the study were conducted under nonpandemic conditions.

XSL•FO
RenderX

Finally, the pilot study length was brief (3 months); while this is sufficient time to detect meaningful BP changes, it is a short time frame in which to produce clinically meaningful weight loss, which may explain the modest weight loss achieved by study participants. Finally, the size of the study sample limited the ability to conduct meaningful subgroup analyses and limited the statistical power for analyses of secondary and tertiary outcomes.

Conclusions

The COVID-19 pandemic has created unexpected opportunities for digital health, with more routine care, including care for chronic conditions, transitioning to remote delivery, greater demand emerging for remotely delivered solutions, and reconsideration of regulations that previously slowed growth and scalability [28]. HTN self-management is an obvious fit for digital health solutions, and the results of this pilot study add to the growing body of evidence that human-supported digital self-management programs can improve outcomes for those with chronic conditions [29]. With proper design, the essential features of supported HTN self-management, including comprehensive counseling for lifestyle changes, HMBP with actionable feedback, and medication taking support, can be effectively translated to a digital format and result in strong program engagement, improved activation, and desirable clinical outcomes for people with HTN. Future research on this program will focus on the sustainability of the clinical outcomes, the robustness of the clinical benefit under increasingly rigorous testing conditions, and research among more diverse populations to promote health equity from new digital health solutions.

References

1. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2020 Update: A Report From the American Heart Association. Circulation 2020 Mar 03;141(9):e139-e596 [FREE Full text] [doi: 10.1161/CIR.0000000000000757] [Medline: 31992061]

2. Kirkland EB, Heincelman M, Bishu KG, Schumann SO, Schreiner A, Axon RN, et al. Trends in Healthcare Expenditures Among US Adults With Hypertension: National Estimates, 2003-2014. J Am Heart Assoc 2018 May 30;7(11):e008731 [FREE Full text] [doi: 10.1161/JAHA.118.008731] [Medline: 29848493]

3. Muntner P, Hardy ST, Fine LJ, Jaeger BC, Wozniak G, Levitan EB, et al. Trends in Blood Pressure Control Among US Adults With Hypertension, 1999-2000 to 2017-2018. JAMA 2020 Sep 22;324(12):1190-1200 [FREE Full text] [doi: 10.1001/jama.2020.14545] [Medline: 32902588]

4. Lebeau J, Cadwallader J, Aubin-Auger I, Mercier A, Pasquet T, Rusch E, et al. The concept and definition of therapeutic inertia in hypertension in primary care: a qualitative systematic review. BMC Fam Pract 2014 Jul 02;15:130 [FREE Full text] [doi: 10.1186/1471-2296-15-130] [Medline: 24989986]

5. Agarwal R, Bills JE, Hecht TJW, Light RP. Role of home blood pressure monitoring in overcoming therapeutic inertia and improving hypertension control: a systematic review and meta-analysis. Hypertension 2011 Jan;57(1):29-38. [doi: 10.1161/HYPERTENSIONAHA.110.160911] [Medline: 21115879]

6. Kerr EA, Zikmund-Fisher BJ, Klamerus ML, Subramanian U, Hogan MM, Hofer TP. The role of clinical uncertainty in treatment decisions for diabetic patients with uncontrolled blood pressure. Ann Intern Med 2008 May 20;148(10):717-727 [FREE Full text] [doi: 10.7326/0003-4819-148-10-200805200-00004] [Medline: 18490685]

7. Wilson PWF, D'Agostino RB, Sullivan L, Parise H, Kannel WB. Overweight and obesity as determinants of cardiovascular risk: the Framingham experience. Arch Intern Med 2002 Sep 09;162(16):1867-1872. [doi: 10.1001/archinte.162.16.1867] [Medline: 12196085]

8. Tucker KL, Sheppard JP, Stevens R, Bosworth HB, Bove A, Bray EP, et al. Self-monitoring of blood pressure in hypertension: A systematic review and individual patient data meta-analysis. PLoS Med 2017 Sep;14(9):e1002389 [FREE Full text] [doi: 10.1371/journal.pmed.1002389] [Medline: 28926573]

9. Margolis KL, Asche SE, Dehmer SP, Bergdall AR, Green BB, Sperl-Hillen JM, et al. Long-term Outcomes of the Effects of Home Blood Pressure Telemonitoring and Pharmacist Management on Blood Pressure Among Adults With Uncontrolled

XSL•FO

RenderX

Hypertension: Follow-up of a Cluster Randomized Clinical Trial. JAMA Netw Open 2018 Sep 07;1(5):e181617 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.1617] [Medline: 30646139]

10. Fisher ND, Fera LE, Dunning JR, Desai S, Matta L, Liquori V, et al. Development of an entirely remote, non-physician led hypertension management program. Clin Cardiol 2019 Feb;42(2):285-291 [FREE Full text] [doi: 10.1002/clc.23141] [Medline: 30582181]

11. Morawski K, Ghazinouri R, Krumme A, Lauffenburger JC, Lu Z, Durfee E, et al. Association of a Smartphone Application With Medication Adherence and Blood Pressure Control: The MedISAFE-BP Randomized Clinical Trial. JAMA Intern Med 2018 Jun 01;178(6):802-809 [FREE Full text] [doi: 10.1001/jamainternmed.2018.0447] [Medline: 29710289]

12. Persell SD, Peprah YA, Lipiszko D, Lee JY, Li JJ, Ciolino JD, et al. Effect of Home Blood Pressure Monitoring via a Smartphone Hypertension Coaching Application or Tracking Application on Adults With Uncontrolled Hypertension: A Randomized Clinical Trial. JAMA Netw Open 2020 Mar 02;3(3):e200255 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.0255] [Medline: 32119093]

13. Lorig KR, Holman HR. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]

14. Shahaj O, Denneny D, Schwappach A, Pearce G, Epiphaniou E, Parke HL, et al. Supporting self-management for people with hypertension: a meta-review of quantitative and qualitative systematic reviews. J Hypertens 2019 Feb;37(2):264-279. [doi: 10.1097/HJH.0000000000001867] [Medline: 30020240]

15. Alessa T, Hawley MS, Hock ES, de Witte L. Smartphone Apps to Support Self-Management of Hypertension: Review and Content Analysis. JMIR Mhealth Uhealth 2019 May 28;7(5):e13645 [FREE Full text] [doi: 10.2196/13645] [Medline: 31140434]

16. Li R, Liang N, Bu F, Hesketh T. The Effectiveness of Self-Management of Hypertension in Adults Using Mobile Health: Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e17776 [FREE Full text] [doi: 10.2196/17776] [Medline: 32217503]

17. Holman H, Lorig K. Patients as partners in managing chronic disease. Partnership is a prerequisite for effective and efficient health care. BMJ 2000 Feb 26;320(7234):526-527 [FREE Full text] [doi: 10.1136/bmj.320.7234.526] [Medline: 10688539]

18. Monitoring Your Blood Pressure at Home. Heart Attack and Stroke Symptoms. URL: https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings/monitoring-your-blood-pressure-at-home [accessed 2021-02-04]

19. Wolf MS, Smith SG, Pandit AU, Condon DM, Curtis LM, Griffith J, et al. Development and Validation of the Consumer Health Activation Index. Med Decis Making 2018 Apr;38(3):334-343 [FREE Full text] [doi: 10.1177/0272989X17753392] [Medline: 29436308]

20. Lorig KR, Sobel DS, Ritter PL, Laurent D, Hobbs M. Effect of a self-management program on patients with chronic disease. Eff Clin Pract 2001;4(6):256-262. [Medline: 11769298]

21. Knobel H, Alonso J, Casado JL, Collazos J, González J, Ruiz I, GEEMA Study Group. Validation of a simplified medication adherence questionnaire in a large cohort of HIV-infected patients: the GEEMA Study. AIDS 2002 Mar 08;16(4):605-613. [doi: 10.1097/00002030-200203080-00012] [Medline: 11873004]

22. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol 2018 May 15;71(19):e127-e248 [FREE Full text] [doi: 10.1016/j.jacc.2017.11.006] [Medline: 29146535]

23. Toro-Ramos T, Kim Y, Wood M, Rajda J, Niejadlik K, Honcz J, et al. Efficacy of a mobile hypertension prevention delivery platform with human coaching. J Hum Hypertens 2017 Dec;31(12):795-800. [doi: 10.1038/jhh.2017.69] [Medline: 28972573]

24. Milani RV, Lavie CJ, Bober RM, Milani AR, Ventura HO. Improving Hypertension Control and Patient Engagement Using Digital Tools. Am J Med 2017 Jan;130(1):14-20. [doi: 10.1016/j.amjmed.2016.07.029] [Medline: 27591179]

25. Guthrie NL, Berman MA, Edwards KL, Appelbaum KJ, Dey S, Carpenter J, et al. Achieving Rapid Blood Pressure Control With Digital Therapeutics: Retrospective Cohort and Machine Learning Study. JMIR Cardio 2019 Mar 12;3(1):e13030 [FREE Full text] [doi: 10.2196/13030] [Medline: 31758792]

26. Mao AY, Chen C, Magana C, Caballero Barajas K, Olayiwola JN. A Mobile Phone-Based Health Coaching Intervention for Weight Loss and Blood Pressure Reduction in a National Payer Population: A Retrospective Study. JMIR Mhealth Uhealth 2017 Jun 08;5(6):e80 [FREE Full text] [doi: 10.2196/mhealth.7591] [Medline: 28596147]

27. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]

28. Bashshur R, Doarn CR, Frenk JM, Kvedar JC, Woolliscroft JO. Telemedicine and the COVID-19 Pandemic, Lessons for the Future. Telemed J E Health 2020 May;26(5):571-573. [doi: 10.1089/tmj.2020.29040.rb] [Medline: 32275485]

29. Wilson-Anumudu F, Quan R, Castro Sweet C, Cerrada C, Juusola J, Turken M, et al. Early Insights From a Digitally Enhanced Diabetes Self-Management Education and Support Program: Single-Arm Nonrandomized Trial. JMIR Diabetes 2021 Feb 22;6(1):e25295 [FREE Full text] [doi: 10.2196/25295] [Medline: 33616533]

Abbreviations

BP: blood pressure
CDCES: certified diabetes care and education specialist
CHAI: Consumer Health Activation Index
DBP: diastolic blood pressure
HMBP: home-measured blood pressure
HTN: hypertension
mHealth: mobile health
RN: registered nurse
SBP: systolic blood pressure
SMAQ: Simplified Medication Adherence Questionnaire

XSL•FO
RenderX

Original Paper

Informal Coping Strategies Among People Who Use Opioids During COVID-19: Thematic Analysis of Reddit Forums

Josh Arshonsky[1*], BA; Noa Krawczyk[1*], PhD; Amanda M Bunting[1*], PhD; David Frank[2*], PhD; Samuel R Friedman[1*], PhD; Marie A Bragg[1*], PhD

[1]Department of Population Health, New York University Grossman School of Medicine, New York, NY, United States

[2]Behavioral Science Training Program in Drug Abuse Research, New York University, New York, NY, United States

[*]all authors contributed equally

Corresponding Author:
Marie A Bragg, PhD
Department of Population Health
New York University Grossman School of Medicine
180 Madison Ave
New York, NY, 10016
United States
Phone: 1 646 501 2717
Fax: 1 212 501 2706
Email: SeedProgramPaperSubmissions@nyulangone.org

Abstract

Background: The COVID-19 pandemic has transformed how people seeking to reduce opioid use access treatment services and navigate efforts to abstain from using opioids. Social distancing policies have drastically reduced access to many forms of social support, but they may have also upended some perceived barriers to reducing or abstaining from opioid use.

Objective: This qualitative study aims to identify informal coping strategies for reducing and abstaining from opioid use among Reddit users who have posted in opioid-related subreddits at the beginning of the COVID-19 pandemic.

Methods: We extracted data from 2 major opioid-related subreddits. Thematic data analysis was used to evaluate subreddit posts dated from March 5 to May 13, 2020, that referenced COVID-19 and opioid use, resulting in a final sample of 300 posts that were coded and analyzed.

Results: Of the 300 subreddit posts, 100 (33.3%) discussed at least 1 type of informal coping strategy. Those strategies included psychological and behavioral coping skills, adoption of healthy habits, and use of substances to manage withdrawal symptoms. In addition, 12 (4%) subreddit posts explicitly mentioned using social distancing as an opportunity for cessation of or reduction in opioid use.

Conclusions: Reddit discussion forums have provided a community for people to share strategies for reducing opioid use and support others during the COVID-19 pandemic. Future research needs to assess the impact of COVID-19 on opioid use behaviors, especially during periods of limited treatment access and isolation, as these can inform future efforts in curbing the opioid epidemic and other substance-related harms.

Introduction

In 2020, with the arrival and devastation of the COVID-19 pandemic, many of the efforts in fighting the preexisting opioid epidemic in the United States were interrupted and the number of opioid-related deaths have since accelerated. Reports from national, state, and local agencies have shown an alarming increase in opioid-and other drug-related mortality—especially from illicitly manufactured fentanyl [1]. In 2021, overdose deaths soared to a record estimate of 93,331, including 69,710 (74.69%) involving opioids [2]. The overall estimate of overdose

deaths in 2020 far eclipsed the 72,000 deaths recorded the previous year, amounting to a 29.4% increase.

COVID-19 has affected nearly every facet of society, and it has seriously burdened people who use opioids (PWUO) and people who use drugs (PWUD) in general. Health experts are concerned how the public health measures responding to the pandemic—particularly quarantine and social distancing rules—may have exacerbated active drug use and drastically changed remission statuses on the population level [3,4]. PWUD are at an increased risk of COVID-19 transmission and infection and are more likely to be hospitalized for severe symptoms [5,6]. Given the severity of this ongoing public health crisis and the likelihood that further pandemics will take place in coming decades [7-13], it is imperative to understand how PWUO have dealt with the physical and psychological changes brought on by the COVID-19 pandemic, particularly in response to stay-at-home orders that limit in-person interactions.

Only a small percentage of PWUO seek formal treatment options, with less than 25% of PWUO estimated to use treatments with evidence-based medications for opioid use disorder (MOUD) [14]. Others may turn to strategies such as social media for health-related information and community support [15]. Online social media platforms enable PWUD to freely share their experiences, give advice, and comment on others' experiences. Studies have shown that for PWUD, online peer support groups provide an open forum for discussion, while minimizing perceived barriers and stigma [16]. In the context of the COVID-19 pandemic, it may serve as a vital source of personal narratives of opioid use and a means to cope with opioid use without formal assistance (ie, informal coping) or seek formal assistance through treatment (ie, formal coping).

Many barriers exist that may prevent PWUO from seeking formal treatment. Social networks, for example, are powerful influences among PWUD that can encourage continued use or serve as a catalyst for return to use [17,18]. Stigma from both health care workers [19] and family members and friends [20,21] may discourage PWUD from disclosing their substance use disorder and receiving treatment. Access to evidence-based MOUD treatments, such as buprenorphine and methadone, are limited by regulations that require in-person examinations, and physician waivers for buprenorphine prescriptions [22] and take-home allowance limits for methadone result in daily in-person visits to clinics [23]. Finally, PWUO may want to discontinue use, but fear the resulting withdrawal symptoms, which can be painful and distressing [24]. The potential length of withdrawal may inhibit PWUO from cessation if they cannot risk losing work or have other obligations that they must fulfill [24].

Although the COVID-19 pandemic has exposed PWUO to new risks (eg, possible triggers due to loneliness, boredom, and isolation, as well as forcing people into withdrawal due to lack of access), behaviors that prevent the spread of COVID-19 may help mitigate barriers to cessation or entry to treatment. The change in environment spurred by social distancing measures may make it easier for some PWUO to avoid influences in their social circle that might encourage them to continue using drugs. Avoiding these influences might ease unwanted pressures while

they make decisions regarding cessation or treatment [17,18]. Still, a large portion of PWUO do not seek formal treatment, making it critical for the clinical and public health field to understand how people outside of the clinical setting cope with opioid use, especially now within the context of the COVID-19 pandemic.

Using data collected from Reddit, this exploratory and descriptive study aims to characterize the types of informal coping strategies people used to reduce or abstain from opioid use during the first wave of the COVID-19 pandemic in 2020.

Methods

Study Design

This study is part of a larger project that sought to understand the experiences of PWUO during the first wave of the COVID-19 pandemic [25,26]. We extracted publicly available content from Reddit, an anonymous and community-based online platform dedicated to public discussions. Anonymous registered members—also called *Redditors*—submit content to the site, including text, images, or links, which are then filtered by popularity as they are voted up or down by other members. Reddit posts are organized by subject and grouped into user-generated forums called *subreddits*. Posts with more up-votes appear toward the top of their subreddit.

Reddit has over 430 million monthly active users, and in 2020, it was ranked the seventh-most popular social media app in the United States [27]. Because of its anonymity, Reddit provides a virtual space to exchange information about stigmatized topics, such as drug use and mental health [28,29]. This, in turn, provides researchers with a unique opportunity to gain, in real time, insights into behaviors and perceptions that can inform health and drug policy. The anonymity of Reddit, however, hinders our ability to retrieve demographic information of Redditors' whose posts we have included in our analysis. However, in many of our sampled subreddit threads, Redditors disclosed their personal history of opioid use and whether, at the moment of posting, they were abstaining from use. So, despite the lack of demographic information on Redditors, the analyses presented in this exploratory study can still inform efforts to advance health and drug policy—albeit to a certain extent—during the COVID-19 pandemic.

As researchers of drug use and social media, we were interested in characterizing the informal ways in which PWUO navigated the disruptions caused by the first wave of COVID-19 in 2020. Though none of the authors subscribe to the subreddits analyzed in this study, 3 of us have used Reddit in the past and are familiar with the platform and communities it coalesces. At each stage of the study, we discussed our methods, findings, and interpretations with individuals who have lived experiences with the subject matter. The New York University School of Medicine Institutional Review Board exempted our study from review. To protect the anonymity of Redditors, quotes were altered slightly to reduce the risk of immediate identification [30,31].

Collecting Reddit Data

In this study, we analyzed posts from the 3 most popular subreddits about opioids: *r/Opiates* and *r/OpiatesRecovery*. We determined their popularity by the number of Redditors subscribed to each subreddit. At the start of data collection, in March 2020, *r/OpiatesRecovery* had more than 20,000 subscribers; *r/Opiates* had over 90,000.

To pull content from each subreddit, we used Python 3.7.4 run on Jupyter Notebook 6.0 using packages *PRAW, pandas*, and *datetime*. On May 13, 2020, we extracted the 1000 most recent posts—including their subsequent comments—from each subreddit. This amounted to a total of 2000 unique subreddit posts with comments, all of which Redditors posted from March 5 to May 13, 2020. After pulling the posts, we stored them in an online, password-protected database.

Reddit and Qualitative Coding Training

We first trained 9 research assistants on how to conduct qualitative coding using Atlas.ti software (ATLAS.ti Scientific Software Development GmbH). Those 9 coders then underwent 4 one-hour training sessions—developed by our team—on medical terms and slang related to opioid use and Reddit culture.

Thematic Analysis

The 9 trained coders then reviewed all 2000 subreddit posts to determine whether they were relevant to COVID-19. To determine relevance, we used the following list of key terms: *coronavirus*, *COVID-19*, *quarantine*, *social distancing*, *shutdown*, *pandemic*, and other related terms referring to the pandemic. To initially identify COVID-related content, coders reviewed only the original subreddit posts and not their subsequent comments. We wanted to avoid situations in which coders reviewed threads of 50 or more comments, only to discover that 1 comment related to COVID-19. Coders identified 300 (15%) original subreddit posts related to COVID-19; we discarded the remaining 1700 (85%) posts because none of them explicitly included our keywords or hinted at any aspect of the pandemic and its effects.

Using a hybrid inductive-deductive approach through Atlas.ti, the 9 research assistants then coded the 300 COVID-19-relevant subreddit posts and comments using a codebook we created prior to data collection [25,26]. The codebook spanned study themes of interest related to the intersection of opioid use, treatment, and COVID-19. Research assistants coded an initial 30 (10%) of the 300 posts. Afterward, we met with them to assess coding discrepancies and add any new or common topics identified from the subreddit posts to our codebook. We used the methodological approach described by McDonald et al [32]. They specify that for research focused on discovering new concepts, codebook development meetings do not aim to achieve a certain threshold of agreement or interrater reliability. Instead, the goal of the codebook development discussions is to generate concepts and themes that enable discovery of novel content for an exploratory study. During the training process, we asked coders to identify any terms that were unfamiliar to them. Then, during group meetings, we defined these terms—which were often slang terms. The final codebook had 44 codes (see Table S1 in Multimedia Appendix 1), and research assistants coded the remaining sample of COVID-19-related posts.

Using solely an inductive approach, the first author analyzed the 300 COVID-19-related posts previously coded by research assistants to identify and characterize Redditors' informal coping strategies during the early weeks of the COVID-19 pandemic in 2020. Based on previous addiction and recovery literature [33], *informal coping strategies* were defined as those that did not involve professionally assisted support services, such as outpatient/inpatient treatment, antirelapse/anticraving medications (eg, naltrexone or buprenorphine/naloxone), mutual help groups (eg, Alcoholics Anonymous [AA], Narcotics Anonymous [NA], Self-Management and Recovery Training [SMART]), and other community-based support where trained staff typically provide services [33]. Of the 44 codes, 16 (36%) were used to filter out posts relevant to informal coping strategies (see Table S2 in Multimedia Appendix 1). The first author discarded subreddit posts that were specific to other research questions (eg, formal treatment services, such as medication-assisted treatment) or if the content did not include or was entirely unrelated to informal coping strategies to maintain remission or begin/continue cessation from opioid use.

After meeting with coauthors and agreeing on the major themes characterizing the informal coping strategies described in posts, the first author scanned nonexamined Reddit content to identify those same strategies and skills. Using the search function feature, he searched through previously unexamined posts for the exact terms or phrases listed in Table 1. After examining the content's relevance to our research aims, the first author either added it to our preexisting sample of content or discarded it. This process resulted in an increase of 12 subreddit posts, making the total sample amount to 100 posts.

Table 1. Frequency of informal coping strategies described by Redditors during the first wave of the COVID-19 pandemic in 2020 (N=100).

Themes	Posts, n (%)
Psychological/behavioral coping skills (eg, psychological and emotional challenges, triggers)	
Mindfulness (eg, reframing personal problems, identifying discontentment)	20 (20)
Meditation	5 (5)
Prayer	2 (2)
Gaming (eg, video games, board games, internet-based games)	12 (12)
Reading books	4 (4)
Listening or making music	4 (4)
Household chores	1 (1)
Watching movies, television, or videos on social media	6 (6)
Calling friends and family	6 (6)
Posting in subreddit or other online forums (eg, encouraging others to post online, disclosing struggle)	5 (5)
Taking up hobbies or learning new skills	9 (9)
Total	74 (74)
Adopting healthy habits	
Exercise	26 (26)
Drinking water	2 (2)
Eating healthy	6 (6)
Sleep hygiene	4 (4)
Total	38 (38)
Using other substances to manage withdrawal symptoms	
Marijuana	9 (9)
Transcutaneous electrical nerve stimulation (TENS) device (ie, device used to stimulate stiff muscles)	1 (1)
Liposomal vitamin C	1 (1)
Kratom	6 (6)
NyQuil	1 (1)
Total	18 (18)

Results

Informal Coping Strategies

Of the 300 subreddit posts included in our sample, 100 (33.3) posts mentioned at least 1 type of informal coping strategy. Table 1 shows the frequency of Redditors' posts that described at least 1 type of informal coping strategy during the pandemic. Examining these messages revealed a myriad of strategies people used for remission and recovery from opioids. In total, there were 4 types of informal coping strategies that Redditors used themselves or suggested to others. The most frequent were psychological and behavioral skills (74/100, 74%), followed by adopting healthy habits (38/100, 38%) and using other substances to manage withdrawal symptoms (18/100, 18%).

Psychological and Behavioral Coping Skills (n=74, 74%)

The first wave of the COVID-19 pandemic upended the ways people worked and lived. Many employers abruptly required employees to work remotely, and others were forced to lay off large numbers of staff. Quite suddenly and unexpectedly, many

Redditors found themselves mostly restricted to their homes—forcing some to reevaluate their work-life balance and forcing others to navigate social distancing and stay-at-home orders without active incomes. Compounding these difficulties was effectively managing reduction in, or cessation of, opioid use while stuck at home.

To confront these particular challenges, Redditors described a variety of psychological and behavioral coping strategies. Although never explicitly stated, practicing mindfulness was a common psychological skill described in our sample of Reddit posts. According to previous work, mindfulness is generally described as a practice to bring awareness of and attention to the present moment without judgement [34]. To be mindful means to observe what is happening—both inside and outside yourself—in the here and now, without evaluating your observations [35]. Redditors described their mindfulness practices in some of the following ways:

> *I try to look for the good things when I'm sober, the chirps of birds, the sound of trees in the wind, and*

the warmth from the sun on my skin always comes back. It's great to finally be able to listen to the world when I'm sober.

Sobriety can be boring but learning to cope with your baseline emotions feels so much better than dealing with the intense highs and lows of opiates. It's not natural and not worth it.

I make videos every day and write down just how fucking awful I feel to go back and really remind myself how bad it is and why I don't ever want to do it again once I'm done.

The subreddits themselves served as a psychological coping strategy in that they provided Redditors with an easily accessible space for communal support. Redditors shared their experiences and strategies—both successful and unsuccessful—with others to encourage their cessation of or reduction in opioid use, supporting and empowering them to push forward and continue making progress:

I made a plan that was damn near [foolproof] where no matter what happened or how bad it got I had a contingency in place to keep me from getting in my car and going and copping. I used that plan more than I thought I would and I'm glad I had it. It included things like calling my brother or sober friends or when I knew I was going to have a tough day I'd just go spend it with my dad or at the animal shelter . . .

In some instances, Redditors posed thoughtful questions to those who disclosed their personal struggles:

You're also doing something important in reaching out and venting on here, those connections are so important, it certainly doesn't sound like dumb bullshit. Sounds like you've got a lot of inner strength from previous experiences, and you can acknowledge what you've learned from them. Maybe part of coping with this is asking what you can learn from this new thing (in a way you are already asking that!), but I know it's really hard.

Redditors exchanged many behavioral strategies to combat boredom and distract themselves to reduce or abstain from opioid use amid sudden COVID-19 circumstances. The most frequently discussed behavioral strategy was gaming via PlayStation and other consoles, as well as via board games and different websites.

Other Redditors distracted themselves at home by picking back up old hobbies and activities or learning new ones:

I'm not working right now so I'm at home by myself for most of the day, until my husband gets home from work. I'm taking advantage of it though. I'm working on a master list of activities around the house I can do that will help me distract from any paranoid thoughts. Right now my list is made up of music-related things (singing, playing ukulele) and learning-related things (continuing an online class I started a month or so ago) and craft-related things

(working on a diamond painting, picking up knitting again).

In some posts, Redditors mentioned watching movies and amusing videos on social media to fight boredom. For 1 Redditor, an effective strategy to redirect their negative thinking of their personal hardship was to watch documentaries that reframe such hardship through context and comparison:

. . . I'm super weird in that when I'm kicking sometimes I like to watch documentaries about really fucked up shit to remind me that like my situation isn't that bad I guess? Lol kinda crazy but it does work for me in terms of just getting my mind out of that horribly dark place . . .

For 1 self-disclosed "sober" Redditor, using marijuana became another important and much-needed tool to combat boredom and maintain remission from opioid use after withdrawal symptoms dissipate:

Pot is a lifesaver for me in sobriety and I'm not even a pot guy really, it helps me with relative boredom that comes with not chasing pills and just sort of allows me to focus on other things. Doesn't do a damn thing for me in active [withdrawals] but post [withdrawals] it's a solid maintenance tool.

Adopting Healthy Habits (n=38, 38%)

On both subreddits, during the early days and weeks of the COVID-19 pandemic in 2020, Redditors also emphasized the crucial role of developing healthier habits—such as exercising, staying hydrated, and eating healthy—to help improve mood and overall well-being. Some Redditors shared their healthy practices for getting more energy throughout the day or for clearing brain fog:

Stay at home but go for a walk or run, drinks and drinks and drinks to stay hydrated, consume funny media, take a nice shower, clean your tub . . . I need to. But get enough sleep exercise and try to eat okay and you'll start thinking clearer and things will get better.

I've been focusing on getting more energy every day. It's been pouring rain where I am for the last few days which sucks [because] I was getting back into jogging and it feels so good!! And I've been doing yoga pretty regularly, too!!

In some instances, Redditors not only described their routines but also disclosed their gradual health effects in real time. For example, 1 Redditor tracked their progress of adopting healthy habits and the impact it had on their sleep:

Wanted to die in week [1], started eating clean in week 2, started exercising/lifting/getting hobbies back in week 3, and sleep started returning to (almost) baseline in week 4. Things are starting to look up again and I can't wait [till] quarantine is over.

Using Other Substances and Devices to Manage Withdrawal Symptoms (n=18, 18%)

Redditors shared information about managing withdrawal symptoms. In some cases, they used definitions of "sobriety" that allowed for certain kinds of substance use (eg, marijuana) but not others:

> I stay clean by hitting my THC vape, and by playing video games or making music. I'm not doing hard drugs because my fiancé is stuck at home with me. She would be very upset if I did.

For clinicians and public health experts, this serves as a valuable reminder that definitions of "sobriety" and acceptable substances may differ across patients.

One Redditor discussed using transcutaneous electrical nerve stimulation (TENS) to help with restless legs and muscle pain during withdrawal. These posts referred to COVID-19 allowing the use of alternative substances and devices to help manage opioid use.

> So I decided to make my first post here, [m]aybe it will help someone! There is this device for muscle problems: Transcutaneous electrical nerve stimulation (TENS) is a method of pain relief involving the use of a mild electrical current. Some people use it to train their muscles without actually training but I don't know if it really works. I personally got it prescribed because of a muscle problem on my neck! It basically gives you a massage and for that purpose its ideal! But I was also able to use a TENS device successfully against my restless legs while withdrawing! If you want to know how it looks just google: Tens/EMS device/unit.

Using Quarantine as an Opportunity for Opioid Cessation or Reduction (n=12, 4%)

Of the 300 subreddit posts included in our sample, 12 (4%) posts mentioned using quarantine as an opportunity for cessation of or reduction in opioid use. Redditors acknowledged that despite the stress caused by COVID-19, the pandemic represented an ideal opportunity to work toward cessation or reduction. The 12 posts referred to quarantine, remote work, and reduction in responsibility as valuable circumstances for their recovery efforts.

> Anyway, as insensitive as it sounds, corona plopped in my lap at a perfect time. It gave me the perfect opportunity to detox and start recovering without any real responsibility.

> You still have options, and the [COVID] shutdown is a great excuse to be able to take a few weeks away from the world and ride out the withdrawals.

> Thankfully, this pandemic has given me an opportunity to begin recovery. But I know I'm one of the lucky ones, as boredom doesn't get to me.

Discussion

Principal Findings

This exploratory study sought to characterize some of the informal coping strategies used by people who were abstaining from use—or simply managing their opioid use as best as they could—while adjusting to the COVID-19 pandemic and remaining at home during the first wave in 2020. To do this, we thematically analyzed a sample of subreddit threads posted on Reddit sometime between March and May 2020. In 12 (4%) of 300 posts, Redditors remarked that the pandemic provided them with an unparalleled opportunity to reduce opioid use. Our analyses revealed 3 overarching themes of informal coping strategies: psychological and behavioral coping skills (74/100, 74%), adopting healthy habits (38/100, 38%), and using substances to manage withdrawal symptoms (18/100, 18%). To help distract themselves from "negative" thinking and the desire to use opioids, Redditors used a variety of strategies. These included activities such as playing video games, taking up old hobbies again, or learning new ones. Redditors also described their adoption of healthy eating habits and exercise and mindfulness practice. In some instances, Redditors disclosed using substances such as marijuana to sway boredom or manage withdrawal symptoms. These exploratory findings may help further our understanding of how the first wave of COVID-19 uniquely affected PWUO, especially people seeking to reduce their opioid use. Ultimately, contributing to this knowledge base will help inform future drug policy and harm reduction efforts responding to future disruptive public health emergencies.

Limitations

Our study, however, is not without limitations. First, the majority of Redditors are younger men living in the United States [36], so the strategies of coping reflected in the subreddit threads may not represent strategies used by the larger community of PWUO. Second, because of the anonymity of Reddit, we had no demographic information or history of drug use for people who posted on both subreddits, so we could not examine the extent to which suggested coping strategies were effective in recovery. Another limitation of our study is that we were not able to verify demographic characteristics and the exact location of the Redditors. It is possible that some of the posts were written by PWUD or their friends or family members. To address this limitation in future studies, researchers could survey patients who are in treatment for OUD and ask them if they have ever used Reddit to discuss opioid use. Third, we did not evaluate the intercoder reliability of our coding frames. We fully believe in the practice and acknowledge its many benefits for qualitative analysis. These benefits include improving the systematicity, communicability, and transparency of the coding process; promoting reflexivity and dialogue among researchers; and helping convince diverse audiences of the trustworthiness of the analysis [37]. With this research, however, we followed the guidelines set by McDonald et al [32]. We did not aim for a certain threshold of agreement. Rather, we sought to generate concepts and themes so that novel content could be discovered and tested in future exploratory studies. Fourth, future studies should return to subreddit threads to examine data during a longer time frame. Our aim with this study, however, was to

understand Redditors' conversations during the first moments of the pandemic. Analyzing their dialogue during the acute time frame enabled us to examine their coping skills during a crisis. Finally, it is unclear whether these Redditors were a highly motivated group that would have succeeded without support from a social media site; it is possible that other social media sites facilitate forums that exacerbate risk factors for opioid use.

Future Implications

These data have several implications for treatment providers, specifically. First, in times of crisis when social interaction is limited, providers can encourage patients to use online forums as a coping strategy and source of communal support. Second, health care workers can listen nonjudgmentally to patients who may describe Reddit as a valuable recovery tool, and be mindful that it represents a safe space for patients who value to anonymity of the forum. Social media and other forums are popular resources for health information, even though some forums may spread misinformation [38]. Studies have shown that people may seek forums and self-help groups to provide emotional support [39-42], particularly given online forums may allow anonymous disclosure without feelings of shame or guilt [43]. Finding emotional support without being stigmatized is especially relevant for PWUD [44] and is an urgent matter, given opioid overdose–related use of health care has increased throughout the pandemic. Future research should examine subgroups of PWUO who may be at heightened risk for reduced access to care during the pandemic [44].

Conclusion

Our current analysis of subreddits supports our previous research suggesting that Reddit facilitates social support that may aid people in their efforts to reduce or abstain from opioid use and improve their physical and mental health [25]. In other research using the same original sample of Reddit content, we found that Redditors sought support and advice from each other on ways to access formal treatment for opioid use during the pandemic [26], whereas a different set of qualitative analyses of these data revealed that some Redditors reported increases or changes in active drug use (D. Frank, PhD, unpublished data, July 2021). Taken together, these findings demonstrate that Reddit is a valuable data source for rapid public health surveillance, especially for emerging public health crises, such as COVID-19.

Multimedia Appendix 1
Final codebook and list of codes used to filter out content for thematic analysis.
[DOCX File , 20 KB - formative_v6i3e32871_app1.docx]

References

1. American Medical Association. Reports of Increases in Opioid- and Other Drug-Related Overdose and Other Concerns during COVID Pandemic. Issue Brief URL: https://www.ama-assn.org/system/files/issue-brief-increases-in-opioid-related-overdose.pdf [accessed 2022-02-17]
2. Ahmad F, Rossen L, Sutton P. Provisional drug overdose death counts. National Center for Health Statistics. URL: https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm#citation [accessed 2022-02-17]
3. Alexander GC, Stoller KB, Haffajee RL, Saloner B. An epidemic in the midst of a pandemic: opioid use disorder and COVID-19. Ann Intern Med 2020 Jul 07;173(1):57-58. [doi: 10.7326/m20-1141]
4. Khatri U, Perrone J. Opioid use disorder and COVID-19: crashing of the crises. J Addict Med 2020;14(4):e6-e7. [doi: 10.1097/adm.0000000000000684]
5. Farhoudian A, Baldacchino A, Clark N, Gerra G, Ekhtiari H, Dom G, et al. COVID-19 and substance use disorders: recommendations to a comprehensive healthcare response. An International Society of Addiction Medicine Practice and Policy Interest Group position paper. Basic Clin Neurosci 2020 May 30;11(2):133-150 [FREE Full text] [doi: 10.32598/bcn.11.covid19.1] [Medline: 32855772]
6. Allen B, El Shahawy O, Rogers E, Hochman S, Khan M, Krawczyk N. Association of substance use disorders and drug overdose with adverse COVID-19 outcomes in New York City: January-October 2020. J Public Health (Oxf) 2021 Sep 22;43(3):462-465 [FREE Full text] [doi: 10.1093/pubmed/fdaa241] [Medline: 33367823]
7. Davis M. The Monster at Our Door: The Global Threat of Avian Flu. New York, NY: The New Press; 2005.
8. Johnson CK, Hitchens PL, Pandit PS, Rushmore J, Evans TS, Young CCW, et al. Global shifts in mammalian population trends reveal key predictors of virus spillover risk. Proc Biol Sci 2020 Apr 08;287(1924):20192736 [FREE Full text] [doi: 10.1098/rspb.2019.2736] [Medline: 32259475]

XSL•FO
RenderX

9.   Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. Cold Spring Harb Perspect Med 2011 Sep 01;1(1):a006841-a006841 [FREE Full text] [doi: 10.1101/cshperspect.a006841] [Medline: 22229120]

10. Wernecke B, Millar DA, Walters M, Ganswindt A, Dziba L, Wright CY. 'Preventing the next pandemic': A 2020 UNEP Frontiers Series Report on zoonotic diseases with reflections for South Africa. S Af. J Sci 2020 Jul 29;116(7/8):1-4 [FREE Full text] [doi: 10.17159/sajs.2020/8531]

11. Wallace R. Big Farms Make Big Flu: Dispatches on Influenza, Agribusiness, and the Nature of Science. New York, NY: NYU Press; 2016.

12. Wallace R, Liebman A, Chaves LF, Wallace R. COVID-19 and circuits of capital. Mon Rev 2020 May 01:1-15. [doi: 10.14452/mr-072-01-2020-05_1]

13. Weiss RA, McMichael AJ. Social and environmental risk factors in the emergence of infectious diseases. Nat Med 2004 Dec;10(12 Suppl):S70-S76 [FREE Full text] [doi: 10.1038/nm1150] [Medline: 15577934]

14. Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: results from the 2018 National Survey on Drug Use and Health (HHS Publication No. PEP19-5068, NSDUH Series H-54). Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2019.

15. Day E, Copello A, Karia M, Roche J, Grewal P, George S, et al. Social network support for individuals receiving opiate substitution treatment and its association with treatment progress. Eur Addict Res 2013;19(4):211-221. [doi: 10.1159/000343827] [Medline: 23391965]

16. De Choudhury M, De S. Mental health discourse on Reddit: self-disclosure, social support, and anonymity. Proc Int AAAI Conf Web Soc Media 2014;8(1):71-80 [FREE Full text]

17. Brown S, Tracy EM, Jun M, Park H, Min MO. Personal network recovery enablers and relapse risks for women with substance dependence. Qual Health Res 2015 Mar 17;25(3):371-385 [FREE Full text] [doi: 10.1177/1049732314551055] [Medline: 25231945]

18. Buchanan AS, Latkin CA. Drug use in the social networks of heroin and cocaine users before and after drug cessation. Drug Alcohol Depend 2008 Aug 01;96(3):286-289 [FREE Full text] [doi: 10.1016/j.drugalcdep.2008.03.008] [Medline: 18455887]

19. McLaughlin D, Long A. An extended literature review of health professionals' perceptions of illicit drugs and their clients who use them. J Psychiatr Ment Health Nurs 1996 Oct;3(5):283-288. [doi: 10.1111/j.1365-2850.1996.tb00127.x] [Medline: 9004621]

20. Corrigan PW, Watson AC, Miller FE. Blame, shame, and contamination: the impact of mental illness and drug dependence stigma on family members. J Fam Psychol 2006 Jun;20(2):239-246. [doi: 10.1037/0893-3200.20.2.239] [Medline: 16756399]

21. Corrigan PW, Kuwabara S, O'Shaughnessy J. The public stigma of mental illness and drug addiction. J Soc Work 2009 Apr 01;9(2):139-147. [doi: 10.1177/1468017308101818]

22. Davis CS, Samuels EA. Continuing increased access to buprenorphine in the United States via telemedicine after COVID-19. Int J Drug Policy 2021 Jul;93:102905 [FREE Full text] [doi: 10.1016/j.drugpo.2020.102905] [Medline: 32811685]

23. Krawczyk N, Fingerhood MI, Agus D. Lessons from COVID 19: are we finally ready to make opioid treatment accessible? J Subst Abuse Treat 2020 Oct;117:108074 [FREE Full text] [doi: 10.1016/j.jsat.2020.108074] [Medline: 32680610]

24. Pergolizzi JV, Raffa RB, Rosenblatt MH. Opioid withdrawal symptoms, a consequence of chronic opioid use and opioid use disorder: current understanding and approaches to management. J Clin Pharm Ther 2020 Oct 27;45(5):892-903. [doi: 10.1111/jcpt.13114] [Medline: 31986228]

25. Bunting AM, Frank D, Arshonsky J, Bragg MA, Friedman SR, Krawczyk N. Socially-supportive norms and mutual aid of people who use opioids: an analysis of Reddit during the initial COVID-19 pandemic. Drug Alcohol Depend 2021 May 01;222:108672. [doi: 10.1016/j.drugalcdep.2021.108672] [Medline: 33757708]

26. Krawczyk N, Bunting AM, Frank D, Arshonsky J, Gu Y, Friedman SR, et al. "How will I get my next week's script?" Reactions of Reddit opioid forum users to changes in treatment access in the early months of the coronavirus pandemic. Int J Drug Policy 2021 Jun;92:103140. [doi: 10.1016/j.drugpo.2021.103140] [Medline: 33558165]

27. Dean B. Reddit Usage and Growth Stats. URL: https://backlinko.com/reddit-users [accessed 2022-02-17]

28. Pandrekar S, Chen X, Gopalkrishna G. Social media based analysis of opioid epidemic using reddit. AMIA Annu Symp Proc 2018:867-876.

29. Sharma R, Wigginton B, Meurk C, Ford P, Gartner CE. Motivations and limitations associated with vaping among people with mental illness: a qualitative analysis of Reddit discussions. Int J Environ Res Public Health 2017;14(1):7 [FREE Full text] [doi: 10.3390/ijerph14010007] [Medline: 28025516]

30. Ayers JW, Caputi TL, Nebeker C, Dredze M. Don't quote me: reverse identification of research participants in social media studies. NPJ Digit Med 2018 Aug 2;1:30 [FREE Full text] [doi: 10.1038/s41746-018-0036-2] [Medline: 31304312]

31. Wilkinson D, Thelwall M. Researching personal information on the public web. Soc Sci Comput Rev 2010 Aug 17;29(4):387-401. [doi: 10.1177/0894439310378979]

32. McDonald N, Schoenebeck S, Forte A. Reliability and inter-rater reliability in qualitative research. Proc CM Hum-Comput Interact 2019 Nov 07;3(CSCW):1-23. [doi: 10.1145/3359174]

XSL•FO
RenderX

33.   Recovery Research Institute. Self-Managed Recovery. URL: https://www.recoveryanswers.org/resource/self-managed-recovery/ [accessed 2022-02-17]

34.   Brown KW, Ryan RM. The benefits of being present: mindfulness and its role in psychological well-being. J Pers Soc Psychol 2003 Apr;84(4):822-848. [doi: 10.1037/0022-3514.84.4.822] [Medline: 12703651]

35.   Shapiro SL, Carlson LE, Astin JA, Freedman B. Mechanisms of mindfulness. J Clin Psychol 2006 Mar;62(3):373-386. [doi: 10.1002/jclp.20237] [Medline: 16385481]

36.   Who Uses YouTube, WhatsApp and Reddit. 2019. URL: https://www.pewresearch.org/internet/chart/who-uses-youtube-whatsapp-and-reddit/ [accessed 2022-02-17]

37.   O'Connor C, Joffe H. Intercoder reliability in qualitative research: debates and practical guidelines. Int J Qual Methods 2020 Jan 22;19:160940691989922. [doi: 10.1177/1609406919899220]

38.   Metzger M, Flanagin A, Medders R. Social and heuristic approaches to credibility evaluation online. J Commun 2010;60(3):413-439. [doi: 10.1111/j.1460-2466.2010.01488.x]

39.   Ridings C, Gefen D. Virtual community attraction: why people hang out online. J Comput Mediat Commun 2004;10(1):JCMC10110. [doi: 10.1111/j.1083-6101.2004.tb00229.x]

40.   Eichhorn K. Soliciting and providing social support over the internet: an investigation of online eating disorder support groups. J Comput Mediat Commun 2008;14(1):67-78. [doi: 10.1111/j.1083-6101.2008.01431.x]

41.   Braithwaite DO, Waldron VR, Finn J. Communication of social support in computer-mediated groups for people with disabilities. Health Commun 1999 Apr;11(2):123-151. [doi: 10.1207/s15327027hc1102_2] [Medline: 16370973]

42.   Finn J. An exploration of helping processes in an online self-help group focusing on issues of disability. Health Soc Work 1999 Aug 01;24(3):220-231. [doi: 10.1093/hsw/24.3.220] [Medline: 10505283]

43.   Kummervold PE, Gammon D, Bergvik S, Johnsen JK, Hasvold T, Rosenvinge JH. Social support in a wired world: use of online mental health forums in Norway. Nord J Psychiatry 2002 Jul 12;56(1):59-65. [doi: 10.1080/08039480252803945] [Medline: 11869468]

44.   Alexander K, Pogorzelska-Maziarz M, Gerolamo A, Hassen N, Kelly EL, Rising KL. The impact of COVID-19 on healthcare delivery for people who use opioids: a scoping review. Subst Abuse Treat Prev Policy 2021 Aug 09;16(1):60 [FREE Full text] [doi: 10.1186/s13011-021-00395-6] [Medline: 34372900]

Abbreviations

MOUD: medication for opioid use disorder
OUD: opioid use disorder
PWUD: people who use drugs
PWUO: people who use opioids

XSL•FO
RenderX

Notes From the Field

Notes From the Field: A Voice-Activated Video Communication System for Nurses to Communicate With Inpatients With COVID-19

Mimi Dunn[1], BA; Adam Landman[1], MD, MS, MIS, MHS; Jennifer Cartright[2], MSN, RN; Anne Bane[3], MSN, RN; Anne Brogan[4], MSN, RN; Caroline Coy[5], MPH; Haipeng Zhang[1], DO, MMSc

[1]Brigham Digital Innovation Hub, Brigham and Women's Hospital, Mass General Brigham, Boston, MA, United States

[2]Brigham and Women's Hospital, Mass General Brigham, Boston, MA, United States

[3]Informatics and Medication Safety, Brigham and Women's Hospital, Mass General Brigham, Boston, MA, United States

[4]Nursing Informatics Team, Brigham and Women's Hospital, Boston, MA, United States

[5]Analytics, Planning, Strategy and Improvement, Brigham and Women's Hospital, Mass General Brigham, Boston, MA, United States

Corresponding Author:
Mimi Dunn, BA
Brigham Digital Innovation Hub
Brigham and Women's Hospital
Mass General Brigham
60 Fenwood Road
Boston, MA, 02115
United States
Phone: 1 8573075995
Email: mdunn15@bwh.harvard.edu

Abstract

With the relaxing of telehealth regulations through the Health Insurance Portability and Accountability Act (HIPAA) waiver notification for Telehealth Remote Communications during the COVID-19 Nationwide Public Health Emergency, our organization had the opportunity to pilot an innovative virtual care solution using a modified consumer-grade voice-activated video communication system (Amazon Echo Show 8) within one inpatient COVID-19 unit. In this brief report, we describe our experiences with implementing the system and general feedback from clinicians, and discuss areas for future development required to enable future scaling of this solution. Our pilot demonstrates the feasibility of deploying a consumer-grade voice assistant device in COVID-19 patient rooms. We found the devices engaging due to the voice technologies and Alexa functionalities for both clinician and patient entertainment. To enable future deployment at scale, enhancements to the Echo Show and data analytics will need to be further explored.

Introduction

The COVID-19 pandemic has resulted in the unprecedented adoption of virtual care tools [1-3]. Ideally, virtualizing patient encounters in the inpatient setting can preserve the sense of human contact and increase capacity for human connection with patients while making interactions feel less rushed [4]. Additionally, inpatient virtual care tools were needed to both reduce use of personal protective equipment (PPE), especially when supplies were constrained, and reduce in-person contact and risk of transmission. Despite the numerous barriers to telemedicine, such as educating staff, cost, reimbursement, access to broadband, and patient digital literacy, telemedicine has flourished during the pandemic, accelerating implementation timelines that may have been much longer without such a catalyst [5].

In April 2020, Mass General Brigham Health System deployed video intercom communication system (VICS) developed in-house that allowed clinical staff to connect over video to a securely configured tablet inside a patient room [4]. This telehealth solution was implemented to reduce PPE usage and maintain human connection at the bedside. Although quite

XSL·FO
RenderX

successful, there were some limitations to VICS—for example, clinical staff required training to use the technology and it was not a hands-free solution, which was an infection control barrier. Solutions in the consumer space, such as Amazon Echo Show (Amazon.com Inc) devices, addressed the hands-free issue and were familiar to staff already but were historically unable to be used in the health care space.

With the relaxing of telehealth regulations through the Health Insurance Portability and Accountability Act (HIPAA) waiver notification for Telehealth Remote Communications during the COVID-19 Nationwide Public Health Emergency [3], the opportunity to explore additional innovative virtual care solutions emerged, including solutions leveraging consumer-grade systems in the clinical setting. One health system found that "the use of consumer products sourced from local vendors is a viable solution for telemedicine systems focusing on speed, reducing costs, and ease of deployment" [6]. In this brief report, we describe our experiences implementing a voice-activated video communication system for clinicians to communicate with inpatients with COVID-19 using consumer-grade hardware (Amazon Echo Shows).

Methods

Overview

This pilot project was conducted at Brigham and Women's Hospital (BWH), a 736-bed, urban academic quaternary care hospital that is a founding member of the Mass General Brigham health care system in Boston, Massachusetts. For the duration of the pilot from September 21, 2020, to November 9, 2020, BWH had a daily average COVID-19 census of 15 patients. In conjunction with nursing leadership, we selected one unit with primarily inpatients with COVID-19 (COVID-19 Special Pathogens Unit) to conduct the pilot project. Selection criteria included a unit that would consistently treat patients with COVID-19 and one that would be adequately equipped to house the pilot devices.

BWH had previously deployed the VICS platform using tablets to improve communication between staff and patients while reducing use of PPE and physical contact time for health care providers treating patients with COVID-19. The configuration of the VICS tablet is locked and has an auto-answer feature enabled so nurses can monitor patients without disturbing them and engage in high-quality 2-way conversations whenever needed without the patient having to take any action.

We used a modified 2-way video communication device (Amazon Echo Show 8) configured to allow drop-in video calls to the patient room to help the nurse conserve PPE by communicating with the patient without having to enter the room numerous times per day. In partnership with a third-party vendor (Aiva Health), a fleet management software system was implemented to enable drop-ins, manage accounts, and provide additional security measures across multiple Echo Show devices. In a similar manner, the Echo Show devices were programmed to allow drop-ins only from the nurse to the patient room. This drop-in function is an instant live connection from a staff device to a patient device, allowing the clinician to see and speak with

the patient and the patient to see and speak with the clinician. Patients are unable to initiate an outbound audio or video call; to contact nurses, they would still need to use the nurse call button in their room.

Amazon Echo Show Device Deployment

A total of 6 patient rooms were included in the pilot. We matched each patient room Echo Show device with a nurse station Echo Show device for a total of 12 devices in the pilot. The Echo Show devices replaced the VICS devices in the 6 patient rooms. The nurse devices were set up at the nurse stations outside each of the rooms (Multimedia Appendix 1). The patient devices were placed on a shelf attached to an intravenous (IV) stand in the patient room (Multimedia Appendix 2). All Echo Show devices were named based on the building name ("Shapiro") and the assigned number of the device. A new Wi-Fi network was used specifically for the Echo Show devices during this pilot to not interfere with clinical care and as a security measure.

We conducted a live training of the Echo Show device with unit staff prior to the go-live date. Along with the live training, tip sheets were provided for both the nurses and the patients. The tip sheet created for the nurses provided an overview, how-to instructions, important notes, and support guidance (Multimedia Appendix 3). Nurses could drop in to a patient room using voice activation by saying, "Alexa, drop in to Shapiro Echo 2" or by selecting "Shapiro Echo Show" on the touch screen of the Echo Show device.

Patients in the pilot rooms were made aware of the pilot via a patient notification sheet affixed to the Echo Show device in the room (Multimedia Appendix 4). A user guide was also created to explain to patients how to operate the Echo Show device in their room (Multimedia Appendix 5).

We did not encourage patients to touch the Echo Show device. Patients were welcome to explore Alexa functionalities—such as asking everyday questions (eg, weather, news, sports) and listening to music—that did not require any personal account information, although we did not market this feature for the pilot.

Periodic retraining occurred during onboarding of new staff and throughout the duration of the pilot. We collected nurse feedback through regular rounding on the units. The nursing team and technology team feedback was then synthesized and collated into this report.

Technical Infrastructure and Security

Amazon has a proprietary infrastructure for IP-based audio and video calls that uses the Session Initiation Protocol. The session is initiated and established entirely on Amazon's network and the 2-way audio/video communication is peer-to-peer between Echo Show devices on the same network, which is end-to-end encrypted and does not go through Amazon's cloud.

Additional security measures were put in place to ensure Echo Show devices met the security standards of our institution including (1) implementing a fleet management system that automatically deleted all data that originated from the Echo Show devices and (2) configuring the Echo Show devices to

connect to a separate Wi-Fi network specifically set up for this pilot to avoid using clinical networks.

Results

Through rounding on the pilot inpatient unit and compiling comments and emails, we gained valuable feedback on the voice-activated video communication system. The nurses found the Echo Show devices easy to use and the privacy controls were well received. They also enjoyed the additional available features on the Echo Show devices that the VICS system did not have.

Discussion

Overview

In addition to the ease of use for video communications using the Alexa drop-in feature, the most positive feedback received was regarding the voice assistant (Alexa) functionality, which was not a feature we advertised. One nurse described how her patient listened to the radio and programming on the Echo using the Alexa functionality and, being from a different country, loved it as it had content she knew and enjoyed. Likewise, nurses enjoyed using the Echo Show at their workstation to listen to the radio using the Alexa functionality, increasing their satisfaction with the system.

Challenges and Opportunities to Improve Future Voice-Activated Video Communication System Enterprise Deployment

During the pilot, there was interest from nursing leadership to deploy these devices in another building following a COVID-19 outbreak. However, this was not feasible due to several challenges with the current consumer version of the Echo Show device (Table 1).

The Echo Show devices support Wi-Fi Protected Access 2 (WPA2) encryption, which only requires a single preshared key to connect to a network. Learning this single preshared key could lead to a system compromise. Future versions of the Echo Show device can address this by enabling access to WPA2 Enterprise, which requires a unique username and password

and a preinstall unique encryption key, thereby providing additional security and enhancing ease of scalability [7].

One of the barriers to implementation was the lack of an official enterprise fleet management solution for the Echo Show devices. A third-party vendor, Aiva Health, created a custom fleet management solution that provided limited enterprise support that was not native to the device. This resulted in the setup of each device becoming a time-consuming activity, taking 1-2 days, as each Echo Show device needed to be configured and tested individually and then brought online by the Amazon and Aiva Health teams with the drop-in feature turned on. If this pilot were to be expanded, additional dedicated team members would be required to manage the implementation.

Due to the HIPAA waiver for telehealth remote communications, this Echo Show device could only be used in COVID-19 patient rooms and had to be removed from the room and unplugged when a patient with COVID-19 was discharged. This made it very difficult for staff to track the devices. No Echo Show devices were lost during the pilot program, but a fleet management solution that included some type of asset tracking system would have improved the experience.

Echo Show device mounting and placement in the rooms and the fixed camera on the device led to additional challenges. The nurses felt that the placement of the Echo Show device on the IV pole was not secure and, depending on the placement of the IV pole, the camera viewing angle from the device could be suboptimal. This led to situations where the Echo Show device was not always facing the patient when the nurse dropped in.

Lastly, there were no built-in data analytics in the Echo Show devices (eg, number of drop-ins, length of drop-in, and number of times voice activation was used), so there was no official way to track device usage. This would have been helpful in understanding the function and usage of the Echo Show device from a quantitative perspective and for measuring the efficacy of this device and comparing it more effectively with usage of the VICS system.

The barriers that previously limited enterprise scaling of voice assistant systems like Amazon Echo Shows are beginning to soften. Third-party vendors, such as Aiva, provide health care–specific fleet management solutions for Alexa devices.

Table 1. Summary of challenges in implementing a voice-activated video communication system and potential solutions.

Challenge	Potential solution
The devices only support Wi-Fi Protected Access 2 (WPA2) networks	Updating future devices to support Extensible Authentication Protocol-Transport Layer Security (EAP-TLS) authentication
Many nurses were accustomed to using the existing VICS[a] solution so they were less interested in adopting the Echo Show device	Provide more live trainings of Echo Show device or completely replace VICS solution with Echo Show device
Device placement on IV[b] pole was suboptimal as it was not sturdy and was not always facing the patient	Creation of custom mounts for the Echo Show device that can affix to the IV pole and pivot
Lack of data analytics made it difficult to understand usage or efficacy of pilot	Ability to extract and summarize key metrics by device, unit, and user, such as number of drop-ins, length of drop-in, and number of times the voice activation was used

[a]VICS: video intercom communication system.

[b]IV: intravenous.

XSL•FO
RenderX

Further work needs to be done on future devices to support wireless network configurations such as WPA2 Enterprise before these devices can truly be considered for "out of the box" enterprise scaling. Currently, without WPA2 Enterprise support, deploying these devices requires close consideration of how to create a safe and secure connectivity plan. For this pilot, we addressed this through the creation of a stand-alone Wi-Fi network specifically for the Echo Shows. We are hopeful that support for these standards will be adopted in future iterations of the Echo Show hardware.

One of the limitations of this paper is that this is a feasibility pilot in a single unit; more formal research needs to be conducted to help understand and inform further implementation and use of voice-activated video communication systems. Further, we did not formally solicit direct patient feedback on the use of the device. Patient experience is an important area to explore in subsequent research. Finally, this implementation was possible under the HIPAA waiver for telehealth remote communications—additional privacy and security review may be necessary for broader health care use in the future [8].

Conclusion

Overall, this pilot demonstrates the feasibility of deploying a consumer-grade voice assistant device in COVID-19 patient rooms. Although there are a variety of technologies that can be used to deliver similar 2-way video communication, we found the Echo Show device engaging; it differentiates itself due to the voice technologies and Alexa functionalities for both clinician and patient entertainment [9]. To enable future deployments at scale, security and privacy enhancements to the Echo Show and data analytics will need to be further explored.

Conflicts of Interest

AL serves on the Abbott Medical Device Cybersecurity Council. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Echo Show device at nurse station.
[PNG File , 991 KB - formative_v6i3e31342_app1.png]

Multimedia Appendix 2
Echo Show device in patient room.
[PNG File , 751 KB - formative_v6i3e31342_app2.png]

Multimedia Appendix 3
Echo Show nurse instructions.
[PNG File , 666 KB - formative_v6i3e31342_app3.png]

Multimedia Appendix 4
Patient notification sheet regarding use of Echo Show device.
[PNG File , 398 KB - formative_v6i3e31342_app4.png]

Multimedia Appendix 5
Echo Show user guide for patients.
[PNG File , 756 KB - formative_v6i3e31342_app5.png]

References

1. Kim E, Morris CA, Klompas M, Zhang H, Landman AB, Eappen S, et al. Coronavirus disease 2019 (COVID-19) screening system utilizing daily symptom attestation helps identify hospital employees who should be tested to protect patients and coworkers. Infect Control Hosp Epidemiol 2021 Nov 10:1-5 [FREE Full text] [doi: 10.1017/ice.2021.461] [Medline: 34753527]
2. Lai L, Sato R, He S, Ouchi K, Leiter R, deLima Thomas J, et al. Usage Patterns of a Web-Based Palliative Care Content Platform (PalliCOVID) During the COVID-19 Pandemic. J Pain Symptom Manage 2020 Oct;60(4):e20-e27 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.07.016] [Medline: 32730951]
3. Lai L, Wittbold KA, Dadabhoy FZ, Sato R, Landman AB, Schwamm LH, et al. Digital triage: Novel strategies for population health management in response to the COVID-19 pandemic. Healthc (Amst) 2020 Dec;8(4):100493 [FREE Full text] [doi: 10.1016/j.hjdsi.2020.100493] [Medline: 33129176]

XSL•FO

RenderX

4.   Schwamm LH, Estrada J, Erskine A, Licurse A. Virtual care: new models of caring for our patients and workforce. Lancet Digit Health 2020 Jun;2(6):e282-e285 [FREE Full text] [doi: 10.1016/S2589-7500(20)30104-7] [Medline: 32382724]
5.   Jalali MS, Landman A, Gordon WJ. Telemedicine, privacy, and information security in the age of COVID-19. J Am Med Inform Assoc 2021 Mar 01;28(3):671-672 [FREE Full text] [doi: 10.1093/jamia/ocaa310] [Medline: 33325533]
6.   Lowe J, Shen S. Off the Shelf: Rapid Deployment of an Emergency Department Telemedicine Platform Using Readily Available Consumer Products. J Emerg Med 2020 Nov;59(5):726-729 [FREE Full text] [doi: 10.1016/j.jemermed.2020.09.026] [Medline: 33097352]
7.   Chai PR, Zhang H, Jambaulikar GD, Boyer EW, Shrestha L, Kitmitto L, et al. An Internet of Things Buttons to Measure and Respond to Restroom Cleanliness in a Hospital Setting: Descriptive Study. J Med Internet Res 2019 Jun 19;21(6):e13588 [FREE Full text] [doi: 10.2196/13588] [Medline: 31219046]
8.   Notification of Enforcement Discretion for Telehealth Remote Communications During the COVID-19 Nationwide Public Health Emergency. US Department of Health and Human Services. URL: https://www.hhs.gov/hipaa/for-professionals/special-topics/emergency-preparedness/notification-enforcement-discretion-telehealth/index.html [accessed 2021-05-19]
9.   Slightam C, Gregory AJ, Hu J, Jacobs J, Gurmessa T, Kimerling R, et al. Patient Perceptions of Video Visits Using Veterans Affairs Telehealth Tablets: Survey Study. J Med Internet Res 2020 Apr 15;22(4):e15682 [FREE Full text] [doi: 10.2196/15682] [Medline: 32293573]

Abbreviations

BWH: Brigham and Women's Hospital
EAP-TLS: Extensible Authentication Protocol-Transport Layer Security
HIPAA: Health Insurance Portability and Accountability Act
IV: intravenous
PPE: personal protective equipment
VICS: video intercom communication system
WPA2: Wi-Fi Protected Access 2

XSL•FO
RenderX

Original Paper

Effects of a National Preventive Intervention Against Potential COVID-19–Related Gambling Problems in Online Gamblers: Self-Report Survey Study

Anders Håkansson[1], PhD; Andreas Sundvall[2], MSc; Axel Lyckberg[2], MSc

[1]Department of Clinical Sciences Lund, Psychiatry, Faculty of Medicine, Lund University, Lund, Sweden
[2]AB Svenska Spel, Visby, Sweden

Corresponding Author:
Anders Håkansson, PhD
Department of Clinical Sciences Lund, Psychiatry
Faculty of Medicine
Lund University
Baravägen 1
Lund, 22100
Sweden
Phone: 46 46175596
Email: anders_c.hakansson@med.lu.se

Abstract

Background: The COVID-19 pandemic has been suspected to increase gambling problems in the population. Several governments introduced COVID-19–specific interventions early with the aim to prevent gambling problems, but their effects have not been evaluated.

Objective: This study aimed to evaluate a Swedish COVID-19–related temporary legislation imposing an automated weekly deposit limit for online casino gambling.

Methods: The study was an anonymous survey sent by a state-owned gambling operator to online gamblers (N=619), among whom 54.0% (n=334) were moderate-risk/problem gamblers who reached the weekly limit on online gambling during the summer of 2020.

Results: Overall, 60.1% (372/619) were aware of having been limited by the COVID-19–related deposit limit, and a minority (145/619, 23.4%) perceived the intervention as fairly bad or very bad. Among those aware of the intervention, 38.7% (144/372) believed the intervention decreased their overall gambling, whereas 7.8% (29/372) believed it rather increased it. However, 82.5% (307/372) reported having gambled at more than one operator after the limit, and the most common gambling type reported to have increased at another operator was online casino (42% among moderate-risk/problem gamblers and 19% among others; *P*<.001). An increase in gambling following the intervention was associated with being a moderate-risk/problem gambler and having negative attitudes toward the intervention.

Conclusions: The weekly deposit limit had relatively high acceptability, but the study highlights the limitations of a single-operator deposit limit, given the high number of gamblers also reporting gambling at other operators and the lower effect in clients with gambling problems.

Introduction

The global spread of SARS-CoV-2, causing the COVID-19 pandemic, is resulting in a number of mental health consequences in the population [1] caused by either the direct effects of the disease or the restrictions imposed on society and the following behavioral changes. One of the public health hazards suggested to be caused by the pandemic is gambling. Increased gambling behavior, at least in subgroups in the population, was feared in the early phases of the pandemic [2]. A priori, problem gambling, including in the group of

XSL•FO
RenderX

individuals who meet the criteria of having a gambling disorder, constitutes a well-known hazard to health and a risk of psychosocial problems [3]. An overview of the research hitherto conducted in the area of gambling behaviors during COVID-19 recently demonstrated the different findings of the studies conducted so far, with several studies demonstrating that gambling practices rather decreased or did not increase, but some reporting an increase among subgroups, including individuals with a higher degree of gambling problems [4].

Thus, altogether, there are hitherto no convincing findings of generally increased gambling due to COVID-19, but there are indications that the potential effects are very unevenly distributed in the population. Sweden in one of the settings where a number of studies have assessed objective gambling activity measures, or self-reported survey data, during COVID-19. Objective measures of gambling activity have shown that gambling at online commercial gambling operators did not increase during the very first phase of the pandemic, when sports events generally were cancelled [5,6]. However, the financial activity of gambling operators demonstrated some likely migration of gambling behavior away from traditional sports betting during the early period of the sports lockdown [7]. In line with the overall picture of a potential gambling increase in subgroups of the population, an early survey study [8] and its similar follow-up [9] demonstrated that 4% and 6%, respectively, stated a self-reported increase in gambling after the onset of the pandemic, with a considerable overrepresentation of individuals with moderate-risk or problem gambling in this group. As an additional data source in this context, treatment uptake at a regional health care facility for gambling disorder patients was not statistically changed by the COVID-19 pandemic [10].

A number of countries took early action with legislation aiming to prevent problem gambling in response to the pandemic. This included gambling bans, bans against gambling advertising, or limits on gambling deposits, for example, in the United Kingdom, Spain, and Latvia [11-13]. A COVID-19–related legislation, effective from July 2, 2020, was also introduced in Sweden, as a response to the debate during the first phase of the COVID-19 pandemic in Sweden, where problem gambling was suggested to be one of the health hazards potentially associated with the societal changes of the pandemic. This legislation was decided by the Swedish parliament after a proposal made by the government, and it included a maximum limit of weekly deposits made to each single operator (at a level of 5000 SEK [around 560 USD] per week) for the following 2 gambling types: (1) online casino services, which are commercialized and offered by a large number of licensed operators within Sweden, and (2) land-based electronic gambling machines, which all belong to a monopoly of a Swedish state-owned gambling operator. An additional feature of the special COVID-19–related legislation was the introduction of a 100 SEK (around 9 USD) limit to the bonuses offered to first-time clients of a gambling operator, which is typically offered by commercial online casino and sports betting services in the country [14], although not by the state-owned operator at the time of this study.

The effect of government policies on problem gambling during the COVID-19 pandemic is largely unknown. A survey study in the general population in Sweden (ie, both gamblers and nongamblers among web panel respondents) examined awareness of the COVID-19–related gambling regulation and the subjective effects of the regulation. A minority of respondents (30%) were aware of the legislation, which was however markedly more well-known among people with at least moderate-risk gambling (56% were aware of the legislation) and in the subgroup of individuals who had ever self-excluded from gambling (78% were aware of the legislation). A very low proportion of respondents reported being influenced by the regulation, with the group split between those reporting an increase or a decrease in their gambling [9]. No study has been able to examine the effects of the legislation in the specific subgroup targeted, that is, people who reach the weekly deposit limit of the legislation. In addition, it is unknown to which extent at-risk gamblers perceive such a limiting regulation as acceptable. For example, it has been suggested that deposit-limiting interventions in high-risk gambling may potentially be perceived as annoying and potentially even stimulate migration to other gambling types with a higher degree of severity [15]. Thus, there is reason to address targeted gamblers' attitudes to this type of COVID-19–specific limit setting and determine whether such attitudes are associated with an increase or decrease in gambling.

This study was carried out by the communication and sustainability division of the state-owned Swedish gambling operator AB Svenska Spel, as a web survey targeting the group affected by the legally imposed limit on deposits. The aim of the study was to assess, in an anonymous sample of online gamblers, who reached the 5000 SEK deposit limit at the Swedish state-owned gambling operator; determine the self-reported effect of the intervention on subsequent gambling behaviors; and evaluate the knowledge about, attitudes to, and experiences of the effects of this intervention.

Methods

Setting

The Swedish gambling market is based on a license system, where gambling operators involved in a number of gambling types can receive a license to operate within Sweden. Operators need to follow Swedish responsible gambling legislation, including an 18-year minimum gambling age and adherence to a national government-based self-exclusion service. The self-exclusion service, administered by the Swedish Gambling Authority, allows for any individual to self-exclude from all licensed gambling (ie, with the exception of physical lotteries, including charity-based lotteries, minor gambling in funfairs, and limit-deposit "restaurant casinos" offering table games in restaurants, which constitute a very minor proportion of the Swedish gambling market). This service, described in previous publications, is a rare example of a nationwide self-exclusion service [16,17]. In Sweden, individuals with online gambling practices constitute the overwhelming majority of patients who seek treatment for a gambling disorder [18]. The prevalence of moderate-risk or problem gambling in the country has been

estimated to be around 1.5%, and among these individuals, around one-third may likely meet the criteria of a disorder [19].

The state-owned operator AB Svenska Spel was traditionally a gambling monopoly and was effective in that role as long as gambling was primarily land-based. Since January 2019, the license-based system allowed for a large number of licensed operators on the Swedish market. AB Svenska Spel has one subdivision that operates in the areas of sports betting, online poker gambling, and online casino gambling, in competition with a large number of commercial operators in these areas. Another subdivision of AB Svenska Spel is a state monopoly responsible for land-based electronic gambling machines and for land-based casinos. The land-based casinos (consisting of 3 state-owned casinos in the 3 major cities of Sweden) were temporarily closed due to the COVID-19 situation from April 2020 to July 2021.

Study Procedure

This is a self-report, electronic, anonymous survey study carried out by AB Svenska Spel. A web link was sent by email to a sample of clients who reached a 5000 SEK gambling limit (at online casino, online bingo, online sports betting, and online poker) after the introduction of the COVID-19–related gambling legislation. The survey was introduced as being sent from the gambling operator AB Svenska Spel to customers who reached this 5000 SEK deposit limit. Invited subjects were informed that the survey was confidential and that its aim was for Svenska Spel to learn more about how to improve its work against problem gambling. An incentive (a 100 SEK gift card of a type that cannot be used for gambling, alcohol, or tobacco products) was offered to respondents.

Individuals who reported being unaware of having been subject to the deposit limit were excluded from the remaining survey. The study was anonymous. Clients reaching the 5000 SEK limit were identified in the client database, where the operator has information about each client's gambling data, but the data collection was carried out by an external consultant (the market survey company Norstat). Thus, the operator and authors were unaware which individuals responded to the survey. Age data were reported on a group level, and given the confidentiality measures, information about gender or geographical location was not collected.

As the study does not involve data that can be directly or indirectly referred to identified individuals, the project does not require ethical permission according to the Swedish ethics in research legislation. Parts of mainly descriptive data from the survey have previously been posted online in Swedish on the operator's home page.

Study Participants

Individuals were selected based on their gambling statistics during weeks 27 to 40 (July 2 to October 4, 2020). Individuals were addressed if they (1) had received at least one automated notification from Svenska Spel because of having deposited 5000 SEK or more during the same week in their joint gambling account involving either online casino, online bingo, online sports betting, or online poker, after the introduction of the COVID-19–related gambling legislation and (2) had lost at least

1000 SEK on online casino, online bingo, or online poker during that week. The latter was specified in order to exclude sports-only bettors who were not affected by the legislation but who may have reached the 5000 SEK limit on their joint sports and casino account of Svenska Spel. This approach identified 4782 individuals, and among them, we selected those who had not actively refused to receive client surveys and who had an email address available. This resulted in a total of 3442 potential participants. According to the funding of the study, recruitment was aimed to stop after reaching around 600 collected responses, and when recruitment was finally halted, a total of 619 responses had been received. The mean age of the participants was 45.0 years (SD 12.3 years), with a median of 45 years (IQR 35-54 years).

Study Variables

The following study variables were collected in the study:

1. Dichotomous question about awareness of the COVID-19 legislation imposing the 5000 SEK deposit limit.
2. Likert-scale response questions about subjective self-reported effects on one's gambling, and opinions about such an intervention and about subjective changes in a number of gambling types following the 5000 SEK limit for online casino gambling and electronic machine gambling.
3. Likert-scale response questions about the acceptability of the intervention; whether the gamblers perceived the government's decision to introduce this intervention as "very bad," "fairly bad," "neither good nor bad," "fairly good," "very good," or "don't know." These responses were dichotomized into "fairly good" or "very good" vs others and "fairly bad" or "very bad" vs others in the statistical analyses.
4. Multiple-choice questions about whether the gambler had gambled more or less on any of the following gambling types: sports betting, horse race betting, land-based restaurant casino, online poker gambling, lotteries/number games, land-based machine gambling with Svenska Spel, land-based machine gambling with other operators, and online casino with other operators; whether no other gambling type involved gambling more/less; or whether the gambler did not know.
5. Dichotomous question about whether one had gambled more or less (or did not know) after the deposit limit at Svenska Spel Sports & Casino, Svenska Spel in general, or other operators.
6. The Problem Gambling Severity Index (PGSI) [20], a validated 9-item survey tool measuring the level of risky and problematic gambling practices. The same assessment instrument has been used in general population surveys on gambling practices in Sweden [19], as well as in the COVID-19–related surveys in Sweden [8,9,21].
7. Age (in years, here reported only on a group level).

Statistical Methods

Data were reported descriptively, and comparisons were made using the chi-square test for categorical data and Student *t* test for continuous data (and Fisher's exact test for cross-tabulation comparisons where one or more of the squares contained less

than five individuals). Finally, for the measure of whether gambling after the intervention had increased or decreased, these 2 measures were studied as outcome measures in 2 separate logistic regression analyses, where age, attitudes toward the intervention, and moderate-risk/problem gambling status were included as independent variables. Here, results were reported using odds ratios with 95% CIs. By default, associations were considered statistically significant at P values below .05.

Results

Levels of Hazardous and Problem Gambling

Full PGSI data were available from a total of 467 individuals, and of these, 25.1% (117/467) were no-risk gamblers, 20.3% (95/467) were low-risk gamblers, 31.3% (146/467) were moderate-risk gamblers, and 23.3% (109/467) were problem gamblers. In addition, another 79 individuals with partly missing answers could be identified as being at least moderate-risk gamblers, based on a PGSI score of 3 or above from available items, resulting in adequate PGSI data from 546 individuals. Thus, in the full sample of 619 individuals, 54% (n=334) were at least moderate-risk gamblers and 34% (n=212) were no-risk or low-risk gamblers, with missing data in the remaining 73 individuals for whom the total score from available PGSI items was 0 to 2. Moderate-risk/problem gamblers were significantly younger than others (43.5 vs 48.3 years, $P<.001$).

Awareness and Acceptability of the Government Regulation

Overall, 60.1% (372/619) were aware of the limit for their deposits by the government's COVID-19–specific gambling regulation. Awareness of the regulation was higher among individuals with moderate-risk/problem gambling than among others (68% vs 48%; $P<.001$), but was unrelated to age (44.5 vs 45.6 years; $P=.29$). In the entire sample, 34.6% (214/619) believed that the regulation was very good, 19.2% (119/619) believed that it was fairly good, and 21.0% (130/619) believed that it was neither good nor bad. Moreover, 9.5% (59/619) stated that it was fairly bad and 13.9% (86/619) stated that it was very bad. Furthermore, 1.8% (11/619) reported that they did not know.

Gambling After the COVID-19 Regulation in Individuals With Awareness of the Regulation

After reaching the 5000 SEK limit, among those reporting being aware of the regulation (n=372), 38.7% (144/372) stated that this had decreased their total gambling, 47.3% (176/372) stated that their gambling remained approximately the same as before, 7.8% (29/372) believed it increased their gambling, and 6.2% (23/372) did not know.

Gambling types perceived to have increased (in other operators) after the intervention were horse race betting (29/372, 7.8%), sports betting (52/372, 14.0%), lotteries and "number games" (18/372, 4.8%), online casino (123/372, 33.1%), online bingo (12/372, 3.2%), online poker (12/372, 3.2%), "restaurant casinos" (7/372, 1.9%), land-based electronic gambling

machines of Svenska Spel (29/372, 7.8%), and other land-based electronic gambling machines (13/372, 3.5%). Additionally, 37.4% (139/372) did not perceive themselves to have increased any gambling type and 9.7% (36/372) did not know.

Moreover, 37.6% (140/372) reported that they decreased their gambling on the Svenska Spel Sports & Casino subdivision, 19.4% (72/372) reported that they decreased their gambling on Svenska Spel overall, 19.4% (72/372) reported that they decreased their gambling at other operators, and 30.6% (114/372) did not know.

We found that 13.4% (50/372) had gambled on casino, poker, or bingo games at only 1 operator since July 1 (since the introduction of the intervention), 55.6% (207/372) had gambled on these games at 2 to 4 operators, 21.2% (79/372) had gambled at 5 to 10 operators, 3.0% (11/372) had gambled at 11 to 15 operators, 2.7% (10/372) had gambled at more than 15 operators, and 4.0% (15/372) did not know. Thus, altogether, 83% had gambled on casino, poker, or bingo games on more than one gambling operator after the deposit limit. Moreover, 40.6% (151/372) reported being limited, following the COVID-19 regulation, from another operator than Svenska Spel during the observation period, 53.2% (198/372) reported not being limited, and 6.2% (23/372) did not know.

Comparison of Moderate-Risk/Problem Gamblers and Others

Among individuals reporting being aware of the COVID-19 regulation, 284 had full data for the PGSI and 327 had full data in the dichotomy of moderate-risk/problem gambling vs no-risk/low-risk gambling. The number of gambling operators reported was strongly associated with moderate-risk/problem gambling ($P<.001$, chi-square linear-by-linear). Among those reporting two or more other gambling operators after reaching the weekly limit at Svenska Spel, moderate-risk/problem gambling was detected in 74% compared with 39% among those reporting 1 operator ($P<.001$).

Differences between moderate-risk/problem gamblers and other gamblers are demonstrated in Table 1. Altogether, moderate-risk/problem gamblers were significantly more likely to report an increase (following the intervention) in Svenska Spel machine gambling, online casino at other operators, and lotteries and number games, and they were markedly less likely to report that they had not increased gambling at other operators, whereas no differences were seen for other gambling types. Moderate-risk/problem gambling was not significantly associated with having decreased Svenska Spel gambling or Svenska Spel Sports & Casino gambling.

Altogether, moderate-risk/problem gamblers were significantly more likely than others to report having increased overall gambling after the introduction of the regulation, and were not more likely to report overall decreased gambling due to the present regulation. They were significantly more likely to perceive the regulation as very good or fairly good, and less likely to perceive it as very bad or fairly bad.

Table 1. Statistical comparisons between moderate-risk/problem gamblers and other gamblers (N=327).

Variable	Moderate-risk/problem gamblers (n=226), n (%)	Other gamblers (n=101), n (%)	*P* value
Postintervention increase			
Svenska Spel machine gambling	22 (9.7)	3 (3.0)	.04
Online casinos, other operators	95 (42.0)	19 (18.8)	<.001
Lotteries and number games	15 (6.6)	0 (0.0)	<.007
Non-Svenska Spel land-based machine gambling	7 (3.1)	4 (4.0)	.74
Restaurant casino gambling	4 (1.8)	1 (1.0)	>.99
Online poker, other operators	8 (3.5)	2 (2.0)	.73
Online bingo, other operators	8 (3.5)	1 (1.0)	.28
Sports betting	36 (15.9)	10 (9.9)	.15
Horse race betting	17 (7.5)	10 (9.9)	.47
Postintervention increase in gambling overall	24 (10.6)	2 (2.0)	.007
No postintervention increase in gambling on other operators	66 (29.2)	59 (58.4)	<.001
Postintervention decrease			
Svenska Spel gambling overall	47 (20.8)	19 (18.8)	.68
Svenska Spel Sports & Casino gambling overall	85 (37.6)	39 (38.6)	.86
Postintervention decrease in gambling overall	88 (38.9)	37 (36.6)	.69
Perceived intervention as very good or fairly good	126 (55.8)	42 (41.6)	.02
Perceived intervention as very bad or fairly bad	53 (23.5)	40 (39.6)	.003

Correlates of Reporting Increased or Decreased Gambling Following the Intervention

Age was unrelated to the reporting of decreased gambling after the regulation (45.9 vs 43.7 years, *P*=.09), but the group reporting increased gambling after the regulation was significantly younger (38.1 vs 45.1 years, *P*<.01). Age was unrelated to the perception of the regulation as very good/fairly good (*P*=.86) or very bad/fairly bad (*P*=.82). Age was also unrelated to a decrease in gambling on Svenska Spel Sports & Casino (*P*=.68), Svenska Spel overall (*P*=.52), and other operators (*P*=.78).

The reporting of increased or decreased gambling following the intervention was associated with attitudes toward the regulation. Among those who perceived the regulation as very good or fairly good, 53.1% (102/192) reported decreased gambling (compared to 23.3% [42/180] among others; *P*<.001) and 4.7% (9/192) reported increased gambling (compared to 11.1% [20/180] among others; *P*=.02). Among those who perceived the regulation as very bad or fairly bad, 18.1% (19/105) reported decreased gambling (compared to 46.8% [125/267] among others; *P*<.001) and 15.2% (16/105) reported increased gambling (compared to 4.9% [13/267] among others; *P*<.001).

In logistic regression analysis (Table 2), a reported increase in gambling following the intervention (n=26) was significantly associated with the opinion that the regulation was very bad or fairly bad and with being a moderate-risk/problem gambler, and was nearly significantly associated with younger age. A reported decrease in gambling following the intervention (n=125) was associated with the opinion that the regulation was very good or fairly good, but not with being a moderate-risk/problem gambler or with age (Table 2).

XSL•FO

RenderX

Table 2. Logistic regression of reported increased gambling or decreased gambling following the intervention (N=327).

Variable	OR[a] (95% CI)	P value
Associations with the reporting of increased gambling following the intervention (n=26)		
Regulation perceived as very bad or fairly bad	4.39 (1.88-10.30)	<.001
Moderate-risk/problem gambler	6.89 (1.53-31.00)	.01
Age	0.96 (0.93-1.00)	.05
Associations with the reporting of decreased gambling following the intervention (n=125)		
Regulation perceived as very good or fairly good	3.43 (2.12-5.54)	<.001
Moderate-risk/problem gambler	1.02 (0.60-1.73)	.94
Age	1.02 (1.00-1.04)	.10

[a]OR: odds ratio.

Discussion

Principal Findings

This study examined self-report of gambling among individuals after online gambling was limited due to a special COVID-19–related gambling regulation imposing a weekly deposit limit on online casino and electronic gambling. The study demonstrated that acceptability of the intervention was generally fair. Although only 6 out of every 10 gamblers reported being aware of the limit introduced by the regulation, a majority perceived the intervention as generally positive. In those who were aware of this, it was considerably more common to perceive that the intervention had decreased their overall gambling than to perceive that it increased their gambling. However, limitations and challenges of the intervention were demonstrated. A large majority gambled at other operators, where 4 out of every 10 gamblers reached the regulated weekly deposit limit. This challenge was more pronounced for online casino, which was by far the gambling type most commonly reported to have increased at other gambling sites after being limited by the present operator. This sample of online gamblers who reached the weekly limit had very high rates of gambling problems, and those who scored positive for at least moderate-risk gambling were considerably more common to report online casino gambling at other sites after being limited, to gamble at more operators, and to perceive themselves to have increased gambling after the introduction of the weekly limit. A self-reported negative effect from the intervention was associated with negative attitudes toward it and with higher gambling problems.

Awareness of the fact that one had reached the deposit limit imposed by the government regulation may seem surprisingly low; only 6 out of every 10 gamblers reported that they were aware. However, consistent with previous general population data assessing knowledge about government policies related to gambling and COVID-19 [9], this awareness was markedly higher in the subgroup with manifest gambling problems. Thereby, it can be assumed that people with highly hazardous gambling patterns constitute a group with particularly high gambling involvement and therefore a higher likelihood of recognizing and understanding the nature of the present intervention. Likewise, it can be argued that the intervention

imposed by the government aims particularly at this group, although it also indicates that awareness, and therefore efficacy, of the intervention may have the potential to increase in gamblers without detected gambling problems.

A subgroup of respondents here reported an increase in gambling subsequent to the intervention (ie, intuitively a negative and unintended effect of the intervention). The reporting of an overall increase in gambling as a consequence of the intervention was rare, but was several times more common in moderate-risk or problem gamblers than in others. Thus, although only a minority reported a markedly deteriorating effect from this intervention, it is clear that individuals with a likely addictive behavior were at higher risk of having experienced such a negative effect in contrast to the intentions of the intervention. It cannot be concluded that an actual increase in gambling was indeed a consequence of the temporary COVID-19 legislation, but again, individuals with higher levels of gambling problems may be at a particular risk of having an unfavorable course in gambling as a reaction to the crisis. In previous research involving prevention or harm reduction methods in problem gambling, it had been stated that subgroups of gamblers may perceive preventive interventions as annoying in a way that would, theoretically, even worsen their gambling practices. While this has been described as unlikely, it has been stated that problem gamblers may be at higher risk of such reactions than other gamblers [15]. Given the very high levels of gambling and detected gambling problems in this study, it cannot be excluded that such negative reactions to an intervention are responsible for one part of the perceived negative effect in some gamblers. Here, it should be kept in mind that a significant proportion of the study sample expressed negative feelings related to this type of intervention, but somewhat contrary to what could be expected, moderate-risk/problem gamblers had somewhat more favorable attitudes toward it.

The present findings may be considered to corroborate previous findings that individuals who report an increase in gambling during these times of COVID-19 have a markedly higher probability of being problem gamblers [8,21]. While the gambling increase was not measured in relation to the pandemic itself, but in relation to a specific regulation in affected gamblers, it further demonstrates that people with hazardous gambling practices are more likely than others to demonstrate

a negative development during the COVID-19 crisis. Likewise, in a survey study among gamblers in the population, those who gambled even in highly restricted gambling types during the lockdown period when sports were generally cancelled, were a group with markedly higher gambling problems [21]. Among US casino gamblers, when casinos closed, a minority migrated to new types of gambling, but this group demonstrated more problematic levels of gambling [22]. Thus, as a general rule from these studies conducted during COVID-19, including the present setting, high levels of gambling during these times of crisis are more likely to be associated with problematic gambling behaviors.

Meanwhile, it should be kept in mind that all studies in the area have not demonstrated an increasing trend in gambling during the pandemic. In Australia, for example, an increase was not seen [23], and in a Canadian study of land-based casino gamblers, during casino closure, a migration toward online gambling was seen primarily among those who had engaged in online casino gambling before, rather than in the full population [24]. Likewise, early observations from the very first phase of the pandemic did not display obvious increases in online gambling activity [5,6]. In line with this, subgroups of gamblers with manifest gambling problems have even reported the COVID-19 period as relieving, due to a decrease in gambling opportunities during some phases [25]. Thus, it is unclear whether the increase in gambling in a subgroup of respondents here and in response to the imposed deposit limit, may be due to pandemic-related effects or the intervention itself, or simply because of a generally increasing course in the gambling practices of these specific individuals.

It can be debated whether the present data support an effect from the type of intervention imposed or whether the nature of the intervention only may invite gamblers to migrate to other gambling types. Again, it is clear from the present data that at least for some gambling types, individuals with risky gambling habits are more likely than others to transfer their gambling to other modalities or gambling types. This corroborates the findings of survey studies in the population, where individuals who increase another gambling type in response to a limit to one gambling type (in this case, the limitation of sports betting during the lockdown period) are more likely to have gambling problems [8,21]. The present type of intervention appears to have at least a relatively high level of acceptability in affected individuals. A majority had a favorable attitude toward it, and a minority claimed to be against it. However, although acceptability was relatively high, only around 1 in 5 respondents believed they had decreased their gambling with other operators due to the intervention. Thus, while the intervention does not technically prevent an individual from continuing to a different gambling operator after reaching the COVID-19–related limit at the first one, the signaling value of the intervention might decrease gambling at other gambling sites. Here, it should again be remembered that the study sample generally involved a very high level of gambling problems, and it can be suspected that the enforcement of a deposit limit may not necessarily invite reflection and a motivational process of change in the individual in the short term.

Three gambling types stood out as being more commonly reported in problem gamblers owing to an increase in gambling practiced in response to the imposed deposit limit. This included machine gambling and lotteries/number games, as well as online casinos at other operators. Online casino was the most commonly cited. The large role of online casinos in problem gamblers in the present setting has been documented previously; for example, it is by far the most common gambling type reported in clients seeking treatment at a regional gambling unit in Sweden [18]. Thus, it is of great interest to conclude that when gamblers are temporarily banned from one online operator service due to the weekly deposit limit, they most commonly turn to this gambling type but at other operator sites. Moreover, again, it confirms the addictive potential of online casino gambling, which is a highly accessible, rapid, and repetitive type of gambling, as the proportion reporting an increase in that gambling type at other operators was markedly more common among individuals with gambling problems.

Our study sample had very high rates of gambling problems. More than half of the full sample represented at least moderate-risk gamblers, and for the further items studied in the subgroup with awareness of the intervention, moderate-risk gamblers made up a large majority. This is further supported by the fact that a majority reported gambling at more operators, even to the extent that 2 of every 5 respondents had experience of reaching the imposed limit at a different operator. It can be argued that the intervention therefore specifically addressed the targeted group, and therefore, from this study, less is known about whether the intervention plays any role in the remaining population (ie, among people with low nonhazardous gambling practices). Over and above the actual effect of the intervention in those facing the limit, it can be argued that an intervention of this nature may have a didactic effect in individuals without current gambling problems, but who may potentially benefit from advice or from the political signal that gambling is a product with addictive potential and has the risk of severe harm during COVID-19.

Limitations

Owing to the confidentiality protocol applied in this survey study, more detailed data, such as gender, geographical location, and previous gambling habits, were not collected. Thus, individual responses could not be linked to any identifying information or to any prior gambling statistics in the databases of Svenska Spel, for whom the identity of respondents remained unknown. While this successfully maintained confidentiality of the respondents, more in-depth data on risk factors could not be detected, and response data clearly rely on the self-report of participants. As in all self-report surveys, the risk of recall bias or other misinformation cannot be disregarded. Moreover, the limited number of participants, which included the first individuals who responded to the invitation to participate, may constitute a risk of bias, as individuals responding first may potentially have a different degree of involvement in these issues and therefore potentially have a different gambling pattern or different opinions than others. Moreover, it should be kept in mind that the population assessed here was recruited from a single gambling operator, and although it operates in diverse areas, such as sports betting, poker gambling, and chance-based

rapid online games (eg, casino slots and bingo), its profile as a state-owned gambling operator may potentially attract a somewhat different group of gamblers than certain other operators in the market.

A strength of the study is that it, quite uniquely, had the ability of addressing one operator's clients with respect to their gambling at other operators, a type of data that cannot easily be obtained from other sources. The present group is also less likely to be examined in detail in larger population surveys, where the group makes up a small minority with extreme gambling patterns, but it could be assessed here. Moreover, the study had the advantage of being able to address a new COVID-19–specific intervention in relatively temporal proximity to the introduction of the intervention.

Conclusions

In a high-level gambling sample exposed to a government-imposed weekly deposit limit aiming to prevent potential COVID-19–related gambling issues, the acceptability of the intervention was relatively high and somewhat higher in problem gamblers, although many exposed individuals were not apparently aware of having been subject to the intervention. The challenges of a single-operator weekly deposit limit were obvious; many of these exposed individuals subsequently gambled at other operators, and in many cases, they gambled at many operators. Self-reported improvement from the intervention was common and self-reported negative effects were rare, but risk gamblers demonstrated a much higher rate of negative effects and, in particular, a high prevalence of online casino gambling at other operators after being limited by the imposed intervention at the operator studied here. Promising attitudes toward this kind of deposit-limit intervention were seen, but the study identified some difficulties, and it may inspire future development of further types of interventions addressing the overall problem of high-risk gamblers. High-level online gamblers constitute a group with great needs, and the gambling operator's own monitoring may identify this group with potential for receiving harm-reducing and therapeutic interventions.

Conflicts of Interest

References

1. Xiong J, Lipsitz O, Nasri F, Lui LM, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: A systematic review. J Affect Disord 2020 Dec 01;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]

2. Håkansson A, Fernández-Aranda F, Menchón J, Potenza M, Jiménez-Murcia S. Gambling During the COVID-19 Crisis – A Cause for Concern. Journal of Addiction Medicine 2020 Aug;14(4):e10-e12. [doi: 10.1097/adm.0000000000000690]

3. Calado F, Griffiths MD. Problem gambling worldwide: An update and systematic review of empirical research (2000-2015). J Behav Addict 2016 Dec;5(4):592-613 [FREE Full text] [doi: 10.1556/2006.5.2016.073] [Medline: 27784180]

4. Brodeur M, Audette-Chapdelaine S, Savard A, Kairouz S. Gambling and the COVID-19 pandemic: A scoping review. Prog Neuropsychopharmacol Biol Psychiatry 2021 Dec 20;111:110389 [FREE Full text] [doi: 10.1016/j.pnpbp.2021.110389] [Medline: 34146652]

5. Auer M, Malischnig D, Griffiths MD. Gambling Before and During the COVID-19 Pandemic Among European Regular Sports Bettors: An Empirical Study Using Behavioral Tracking Data. Int J Ment Health Addict 2020 May 29:1-8 [FREE Full text] [doi: 10.1007/s11469-020-00327-8] [Medline: 32837423]

6. Lindner P, Forsström D, Jonsson J, Berman AH, Carlbring P. Transitioning Between Online Gambling Modalities and Decrease in Total Gambling Activity, but No Indication of Increase in Problematic Online Gambling Intensity During the First Phase of the COVID-19 Outbreak in Sweden: A Time Series Forecast Study. Front Public Health 2020 Sep 29;8:554542 [FREE Full text] [doi: 10.3389/fpubh.2020.554542] [Medline: 33117770]

7. Håkansson A. Effects on Gambling Activity From Coronavirus Disease 2019-An Analysis of Revenue-Based Taxation of Online- and Land-Based Gambling Operators During the Pandemic. Front Psychiatry 2020 Dec 17;11:611939 [FREE Full text] [doi: 10.3389/fpsyt.2020.611939] [Medline: 33391059]

8. Håkansson A. Changes in Gambling Behavior during the COVID-19 Pandemic-A Web Survey Study in Sweden. Int J Environ Res Public Health 2020 Jun 05;17(11):4013 [FREE Full text] [doi: 10.3390/ijerph17114013] [Medline: 32516880]

9. Håkansson A, Widinghoff C. Changes of Gambling Patterns during COVID-19 in Sweden, and Potential for Preventive Policy Changes. A Second Look Nine Months into the Pandemic. Int J Environ Res Public Health 2021 Feb 27;18(5):2342 [FREE Full text] [doi: 10.3390/ijerph18052342] [Medline: 33673575]

10. Håkansson A, Åkesson G, Grudet C, Broman N. No Apparent Increase in Treatment Uptake for Gambling Disorder during Ten Months of the COVID-19 Pandemic-Analysis of a Regional Specialized Treatment Unit in Sweden. Int J Environ Res Public Health 2021 Feb 17;18(4):1918 [FREE Full text] [doi: 10.3390/ijerph18041918] [Medline: 33671210]

11. Sweden to limit online betting during coronavirus outbreak. Reuters. 2020. URL: https://tinyurl.com/aejnksyp [accessed 2021-08-20]

XSL•FO

RenderX

12.   Spain orders "social shield" to fast track gambling advertising window. SBC News. 2020. URL: https://tinyurl.com/2p8kbzhj [accessed 2021-08-20]

13.   Coronavirus: gambling firms urged to impose betting cap of £50 a day. The Guardian. 2020. URL: https://tinyurl.com/mr42s2vz [accessed 2021-08-20]

14.   Government introduces temporary responsible gambling measures. Swedish Government. URL: https://tinyurl.com/2s4c4txz [accessed 2021-08-20]

15.   Ivanova E, Rafi J, Lindner P, Carlbring P. Experiences of responsible gambling tools among non-problem gamblers: A survey of active customers of an online gambling platform. Addict Behav Rep 2019 Jun;9:100161 [FREE Full text] [doi: 10.1016/j.abrep.2019.100161] [Medline: 31193727]

16.   Håkansson A, Henzel V. Who chooses to enroll in a new national gambling self-exclusion system? A general population survey in Sweden. Harm Reduct J 2020 Oct 21;17(1):82 [FREE Full text] [doi: 10.1186/s12954-020-00423-x] [Medline: 33087113]

17.   Håkansson A, Widinghoff C. Gambling Despite Nationwide Self-Exclusion-A Survey in Online Gamblers in Sweden. Front Psychiatry 2020 Dec 2;11:599967 [FREE Full text] [doi: 10.3389/fpsyt.2020.599967] [Medline: 33343428]

18.   Håkansson A, Mårdhed E, Zaar M. Who Seeks Treatment When Medicine Opens the Door to Pathological Gambling Patients-Psychiatric Comorbidity and Heavy Predominance of Online Gambling. Front Psychiatry 2017 Nov 29;8:255 [FREE Full text] [doi: 10.3389/fpsyt.2017.00255] [Medline: 29238309]

19.   Abbott M, Romild U, Volberg R. The prevalence, incidence, and gender and age-specific incidence of problem gambling: results of the Swedish longitudinal gambling study (Swelogs). Addiction 2018 Apr 24;113(4):699-707. [doi: 10.1111/add.14083] [Medline: 29105942]

20.   Wynne H, Ferris J. The Canadian Problem Gambling Index: Final Report. Canadian Consortium for Gambling Research. 2001. URL: https://tinyurl.com/h4jwaw3m [accessed 2022-01-28]

21.   Håkansson A. Impact of COVID-19 on Online Gambling - A General Population Survey During the Pandemic. Front Psychol 2020 Sep 25;11:568543 [FREE Full text] [doi: 10.3389/fpsyg.2020.568543] [Medline: 33101137]

22.   Xuereb S, Kim H, Clark L, Wohl M. Substitution behaviors among people who gamble during COVID-19 precipitated casino closures. International Gambling Studies 2021 Apr 05;21(3):411-425 [FREE Full text] [doi: 10.1080/14459795.2021.1903062]

23.   Gainsbury S, Swanton T, Burgess M, Blaszczynski A. Impacts of the COVID-19 Shutdown on Gambling Patterns in Australia: Consideration of Problem Gambling and Psychological Distress. Journal of Addiction Medicine 2021 Dec;15(6):468-476. [doi: 10.1097/adm.0000000000000793]

24.   Price A. Online Gambling in the Midst of COVID-19: A Nexus of Mental Health Concerns, Substance Use and Financial Stress. Int J Ment Health Addict 2020 Jul 13:1-18 [FREE Full text] [doi: 10.1007/s11469-020-00366-1] [Medline: 32837444]

25.   Donati MA, Cabrini S, Capitanucci D, Primi C, Smaniotto R, Avanzi M, et al. Being a Gambler during the COVID-19 Pandemic: A Study with Italian Patients and the Effects of Reduced Exposition. Int J Environ Res Public Health 2021 Jan 07;18(2):424 [FREE Full text] [doi: 10.3390/ijerph18020424] [Medline: 33430353]

Abbreviations

PGSI: Problem Gambling Severity Index

XSL•FO

RenderX

<u>Original Paper</u>

The Mental Health Impact of the COVID-19 Pandemic Among Physicians, Nurses, and Other Health Care Providers in Alberta: Cross-sectional Survey

Hany El Gindi[1], MD, MSc-CCM, MRCP; Reham Shalaby[2], MD; April Gusnowski[3], BA; Wesley Vuong[3], MPH; Shireen Surood[3], PhD; Marianne Hrabok[4], PhD; Andrew J Greenshaw[2], PhD; Vincent Agyapong[2], MD, PhD

[1]Critical Care Medicine Department, King Abdul-Aziz Hospital, Jeddah, Saudi Arabia

[2]Department of Psychiatry, University of Alberta, Edmonton, AB, Canada

[3]Addiction and Mental Health, Alberta Health Services, Edmonton, AB, Canada

[4]Department of Psychiatry, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Corresponding Author:
Vincent Agyapong, MD, PhD
Department of Psychiatry
University of Alberta
1E1 Walter Mackenzie Health Sciences Centre
8440 112 St NW
Edmonton, AB, T6G 2B7
Canada
Phone: 1 780 215 7771
Email: agyapong@ualberta.ca

Abstract

Background: During the COVID-19 pandemic, threats to mental health, psychological safety, and well-being are evident, particularly among the first responders and the health care staff.

Objective: This study aims to examine the prevalence and potential predictors of the likely stress, generalized anxiety disorder, and major depressive disorder among health care workers (HCWs).

Methods: A cross-sectional survey was used through a survey link sent to gather demographic information and responses on several self-report scales, including the Perceived Stress Scale, the Generalized Anxiety Disorder 7-item scale, and the Patient Health Questionnaire-9 among HCWs enrolled in the Text4Hope program.

Results: The result from this study suggests that during the COVID-19 pandemic, HCWs reported a high likelihood of moderate-to-high perceived stress (n=840, 81.2%), moderate-to-severe anxiety (n=369, 38.6%), and depression (n=317, 32.7%) symptoms. Nurses and other HCWs were significantly more likely to report depressive symptoms compared to physicians (F(2, 159.47)=15.89, 95% CI –5.05 to –2.04). Younger age groups of HCWs (≤30 years) were more prone to report likely stress, anxiety, and depressive symptoms compared to HCWs 41-50 and >50 years old (odds ratio [OR] 1.82-3.03). Similarly, females and those who reported a lack of social support (separated/divorced and single) among HCWs had a higher likelihood to report likely stress and depressive symptoms, respectively (OR 1.8 and 1.6, respectively).

Conclusions: This cross-sectional study explored a high level of mental health burdens during the COVID-19 pandemic among HCWs in Alberta. Levels of psychological symptoms were more noticeable in the female gender and the nursing profession.

Introduction

Background

During pandemics, an exponential increase in the demand for health care services takes place. Several factors contribute to increased physical and mental health strain for health care workers (HCWs). This can include long work shifts, a lack of personal protective equipment, and limited resources to care for patients [1,2]. With a lack of PPE, HCWs can feel unprepared to deal with unknown viruses and bacteria. In addition, with pandemic demand–induced limits on resources, such as ventilators or general medical supplies, it can be difficult to care for patients.

In November 2019, a novel coronavirus disease caused by SARS-CoV-2, called COVID-19, was first reported in Wuhan, China. The disease rapidly spread throughout China and prevailed worldwide, resulting in a global health emergency [1]. On March 11, 2020, the World Health Organization declared the COVID-19 outbreak as a global pandemic [2].

Alberta Health Services (AHS) defines health care professionals/providers as "individuals that work in the health field and can include doctors, nurses (RNs, LPNs), dentists, psychologists, physiotherapists, pharmacists, and dieticians, etc" [3]. The nursing staff constitutes the largest health profession in Canada, followed by physicians and other regulated HCWs [4]. As of July 2020, the Canadian Institute for Health Information (CIHI) reported that 19.4% of all confirmed infections of COVID-19 cases were among HCWs in Canada, and in the province of Alberta, it was 8.8% [5]. HCWs are exposed to mental health stresses due to the nature of their work that usually involves trauma and vulnerability [6].

During the 2003 SARS outbreak, HCWs expressed immediate psychological distress manifested as fear and anxiety that exhibited a relative decrease during the early phases of the epidemic [1]. However, depression and posttraumatic stress symptoms emerged later during the epidemic and lasted for longer periods, affecting the long-term mental well-being of HCWs [1]. A greater psychological impact was coupled with the higher risk of exposure to the virus, especially among frontliners, who usually face both heavy workloads and a higher risk of infection [7,8]. One year after the SARS outbreak, HCWs who were in close contact with infected patients or virus material had elevated levels of stress, depression, and anxiety [9].

One study examining the psychological effects of COVID-19 on HCWs in Italy found that depression and posttraumatic stress symptoms are higher in HCWs caring for patients in COVID-19 wards compared to other HCWs caring for patients in other units [10].

Additionally, emotional impacts in terms of contagion fear and infecting loved ones, uncertainty, and stigma were documented among health care staff [11], leading to isolation from their families, changing routines, and a narrowing down of their social support network [12]. During the pandemic of COVID-19, health care professionals declared 5 main requests from their institutions [10]: "hear me, protect me, prepare me, support me, and care for me" [7]. Burnout is also an ongoing problem among physicians and other HCWs during the COVID-19 pandemic. It is characterized by "emotional exhaustion, depersonalization, and a feeling of low personal accomplishment" and attributed to many factors, including heavy workloads, that may exacerbate burnout and negatively impact the overall productivity of the health care system [13,14]. HCWs are predisposed to "moral injury" during the current pandemic, a term that can be defined as *the psychological distress that results from actions, or the lack of them, which violate someone's moral or ethical code*. Although it is not a mental health condition, people with moral injury are more likely to experience negative thoughts about themselves or others, together with intense feelings of shame, guilt, or disgust [15].

During the COVID-19 pandemic, several studies have examined the impacts of the pandemic upon the psychological health and mental well-being of the general population. A number of studies have examined such impacts among HCWs, with many interesting findings; for example, the frontliners and those who have experienced physical symptoms, such as headache, throat pain, and lethargy are at more risk of developing stress, anxiety, and depressive symptoms during the pandemic compared to the other comparative groups [16,17].

In a systematic review assessing the impact of COVID-19 on HCWs' mental health, the authors found that the prevalence of anxiety was estimated between 9 and 90% (median 24%) and depression between 5 and 51% (median 21%) [18]. In another systematic review and meta-analysis, Pappa et al [19] examined 13 research studies to compute the prevalence of anxiety, depression, and insomnia during the current pandemic. The authors reported 23.2% for anxiety and 22.8% for depression. Additionally, discrepancies of the prevalence or severity of such symptoms were reported among different demographic and occupational groups, such as in nurses, women, and frontliners who were usually reported to have more severe symptoms compared to other HCWs [19,20].

Aim and Objectives

This study aims to examine the psychological impacts of the COVID-19 pandemic among different groups of HCWs who were enrolled in the Text4Hope program.

The primary objectives are:

- Studying the demographic characteristics and the prevalence and mean scores of perceived stress, likely major depressive disorder (MDD), and likely generalized anxiety disorder (GAD) among HCWs.
- Studying the predictors of developing likely stress, MDD, and GAD among HCWs.

Methods

Study Design

A cross-sectional survey was used to explore mean differences in perceived stress, anxiety, and depression symptom scores of HCWs enrolled in Text4Hope.

Ethics

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committee on human experimentation and with the Declaration of Helsinki of 1975, as revised in 2008. Participant consent was implied by submission of the subscribers' survey responses. The University of Alberta Health Research Ethics Board provided ethics approval for this research (Pro00086163).

Recruitment

The study recruitment procedures and sample size estimations are described in a published study protocol [21]. The study participants are subscribers to the Text4Hope program, a daily supportive text message service, launched by the AHS on March 23, 2020, to help Albertans cope with the mental health effects of the COVID-19 pandemic. An online survey link was sent to subscribers who were enrolled during March 23 to May 2, 2020. Subscribers receive free daily supportive text messages, which are cognitive behavioral therapy–based messages created by a team of mental health professionals [22]. In addition to demographic information, clinical characteristics were assessed using self-report scales, including the Perceived Stress Scale (PSS-10; for moderate-to-high stress, PSS≥14) [23], the Generalized Anxiety Disorder 7-item (GAD-7) scale (for likely GAD, GAD-7≥10) [24], and the Patient Health Questionnaire-9 (PHQ-9; for likely MDD, PHQ-9≥10) [25]. The PSS-10 is a validated 10-item questionnaire used to assess the self-reported level of stress in the previous 1 month by assessing thoughts and feelings. Each item on the scale is scored from 0 (never) to 5 (very often). Higher scores on the scale denote higher levels of stress. The GAD-7 is a validated 7-item questionnaire used to assess the self-reported levels of anxiety in respondents in the 2 weeks prior to assessment. It is based on the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision* (DSM-IV-TR) symptoms of anxiety. Each item on the scale is scored from 0 (not at all) to 4 (nearly every day). The PHQ-9 is a 9-item validated instrument used to diagnose and measure the severity of depression in general medical and mental health settings; it is the major depression module of the full Patient Health Questionnaire (PHQ). Each of the 9 items on the questionnaire is scored from 0 (not at all) to 3 (nearly every day). It may be used to plan and monitor treatment of depression. Participant consent was implied by submission of the subscribers' survey responses.

Data Analysis

Data analysis was undertaken using the IBM Statistical Package for Social Sciences (SPSS) Statistics for Windows, version 25 [26]. Descriptive analysis illustrated the differences between the 3 categories of HCWs (physicians, nurses, and other) by their sociodemographic characteristics (ie, gender, age, ethnicity, education, relationship status, housing status, and self-isolation/quarantine status) and examined the prevalence of clinical characteristics (ie, moderate/high stress, likely GAD, likely MDD) among the 3 HCW groups.

One-way ANOVA with 2-tailed significance ($P<.05$) was performed to assess the statistical differences of mean scores on the PSS-10, GAD-7, and PHQ-9 among HCW groups. For variables that did not violate the assumptions of homogeneity of variance in the mean scores on the ANOVA test, we performed a Tukey posthoc test. For variables that violated the homogeneity of variance assumption, we used the nonparametric Welch F test and Games-Howell posthoc tests.

To examine potential predictors of the self-reported clinically meaningful symptoms of moderate-to-high perceived stress, likely GAD, and likely MDD, we entered all demographic predictors and self-isolation, along with HCW type, into a multivariate logistic regression model. Correlation analysis was performed before logistic regression analysis to rule out strong correlations among predictor variables. Odds ratios (ORs) from the binary logistic regression analysis were examined to determine the association between HCW type and the likelihood of respondents self-reporting symptoms of moderate-to-high stress, likely GAD, and likely MDD, controlling for the other variables in the model.

Results

Participant Characteristics

Of the 8267 subscribers who responded to the Text4Hope online survey in the first 6 weeks, 1414 (17.1%) self-identified as HCWs, while 1096 (13.3%) provided their specific occupation type. Of these 1096, 63 (5.7%) were physicians, 355 (32.4%) nurses, and 678 (61.9%) other HCWs (eg, occupational therapist, psychologist, dietitian, pharmacist, and first responder). The demographic and clinical characteristics of the respondents are shown in Table 1.

As presented in Table 1, most of the 1096 respondents were female (n=1006, 91.8%), aged 31-40 years (n=330, 30.1%), and Caucasian (n=919, 83.9%); had postsecondary education (n=1066, 97.3%); were married, cohabiting, or partnered (n=852, 77.7%); and owned their own home (n=836, 76.3%). Except for a few variables, nurses represented the highest percentage of these responses.

Table 1. Distribution of the demographic and clinical characteristics by HCW[a] type.

Variables	Physicians (N=63)	Nurses (N=355)	Other (N=678)	Total (N=1096)[b]
Gender, n (%)				
Male	17 (27.0)	10 (2.8)	55 (8.1)	82 (7.5)
Female	46 (73.0)	343 (96.9)	617 (91.1)	1006 (92.0)
Other	0	1 (0.3)	5 (0.7)	6 (0.5)
Age (years), n (%)				
≤30	13 (21.0)	5 (16.8)	90 (13.5)	162 (15.0)
31-40	9 (14.5)	107 (30.4)	214 (32.2)	330 (30.6)
41-50	19 (30.6)	78 (22.2)	193 (29.0)	290 (26.9)
>50	21 (33.9)	108 (30.7)	168 (25.3)	297 (27.5)
Ethnicity, n (%)				
Caucasian	48 (76.2)	309 (87.0)	562 (83.0)	919 (83.9)
Indigenous	1 (1.6)	6 (1.7)	24 (3.5)	31 (2.8)
Asian	10 (15.9)	15 (4.2)	44 (6.5)	69 (6.3)
Other	4 (6.3)	25 (7.0)	47 (6.9)	76 (6.9)
Education, n (%)				
Less than high school diploma	0	1 (0.3)	2 (0.3)	3 (0.3)
High school diploma	0	1 (0.3)	16 (2.4)	17 (1.6)
Postsecondary education	63 (100.0)	351 (99.2)	652 (96.4)	1066 (97.5)
Other education	0	1 (0.3)	6 (0.9)	7 (0.6)
Relationship status, n (%)				
Married/cohabiting/partnered	51 (81.0)	291 (82.0)	510 (75.3)	852 (77.8)
Separated/divorced	3 (4.8)	24 (6.8)	61 (9.0)	88 (8.0)
Widowed	0	3 (0.8)	3 (0.4)	6 (0.5)
Single	9 (14.3)	37 (10.4)	98 (14.5)	144 (13.2)
Other	0	0	5 (0.7)	5 (0.5)
Housing status, n (%)				
Own home	49 (77.8)	287 (80.8)	500 (73.9)	836 (76.3)
Living with family	4 (6.3)	12 (3.4)	24 (3.5)	40 (3.7)
Renting	10 (15.9)	55 (15.5)	147 (21.7)	212 (19.4)
Other	0	1 (0.3)	6 (0.9)	7 (0.6)
Self-isolate/quarantine, n (%)				
No	41 (65.1)	273 (77.1)	540 (80.1)	854 (78.3)
Yes	22 (34.9)	81 (22.9)	134 (19.9)	237 (21.7)
Symptoms, n (%)[c]				
Perceived stress (moderate to high[d])	41 (68.3)	279 (82.5)	520 (81.6)	840 (81.2)
Likely GAD[e]	20 (36.4)	116 (37.4)	233 (39.4)	369 (38.6)
Likely MDD[f]	9 (16.1)	99 (31.3)	209 (35.1)	317 (32.7)
Clinical condition severity, mean (SD)				
PSS-10[g]	18.07 (7.25)	19.25 (6.27)	19.35 (6.28)	19.25 (6.34)
ANOVA $F_{(2, 1032)}$=1.13, P=.32				

Variables	Physicians (N=63)	Nurses (N=355)	Other (N=678)	Total (N=1096)[b]
GAD-7[h]	4.46 (4.36)	7.42 (5.17)	8.01 (5.72)	7.61 (5.53)
ANOVA $F_{(2, 953)}$=0.80, P=.45				
PHQ-9[i]	7.58 (5.25)	8.18 (5.32)	8.46 (5.58)	8.32 (5.48)
Welch statistics $F(2, 159.47)$=15.89, P<.001				

[a]HCW: health care worker.

[b]Data were reported according to the provided responses. The total percentages may not add up to 100%.

[c]Multiple responses were possible.

[d]Moderate-to-high stress was defined as PSS-10 score of ≥14.

[e]Likely GAD was defined as a GAD-7 score of ≥10.

[f]Likely MDD was defined as a PHQ-9 score of ≥10.

[g]PSS-10: Perceived Stress Scale.

[h]GAD-7: Generalized Anxiety Disorder 7-item.

[i]PHQ-9: Patient Health Questionnaire-9.

Among the 1096 respondents, 237 (21.7%) reported having self-isolated/quarantined during the pandemic time. Among those who reported their clinical symptoms, most respondents who self-reported moderate-to-high perceived stress were nurses (279/338, 82.5%), while those who self-reported likely GAD and likely MDD were other HCWs (233/591, 39.4%) and (209/596, 35.1%), respectively.

The mean scores for all HCWs were 19.25 (SD 6.34, n=1035) on the PSS-10 scale, 8.32 (SD 5.48, n=956) on the GAD-7 scale, and 7.61 (SD 5.53, n=968) on the PHQ-9 scale.

One-way ANOVA revealed a significant difference among the 3 groups of HCWs only in terms of their PHQ-9 mean scores. The Welch test and Games-Howell posthoc tests showed that physicians (mean=4.46, SD 4.36) scored significantly lower on the PHQ-9 scale compared to nurses (mean=7.42, SD 5.17, 95% CI –4.51 to –1.4, P<.01) and to other HCWs (mean=8.01, SD 5.72, 95% CI –5.05 to –2.04, P<.001).

Logistic Regression

Spearman correlation analysis revealed no high collinearity among the suggested variables (r_s<0.4), so all the variables were entered into the multivariate regression model.

Table 2 presents data indicating that for moderate-to-high likely stress, the model containing the 8 predictors was statistically significant (χ^2_{21}=68.76, P<.001). The model explained between 6.6% (Cox and Snell R^2) and 10.6% (Nagelkerke R^2) of the variance and correctly classified 80.8% of all cases. The type of HCWs did not significantly contribute to the likely stress model among all the HCWs. Controlling for all other factors in the model, age categories made a unique statistical contribution (Wald=34.24, P<.001) to the likelihood that a respondent presented with moderate-to-high stress.

Table 2. Logistic regression predicting likelihood for the respondents among HCWs[a] to present with moderate-to-high stress.

Variables	B	SE	Wald	df	P value[b]	OR[c] (95% CI)
Gender						
Male	__[d]	—	5.451	2	.07	—
Female	0.683	0.293	5.442	1	.02	1.980 (1.115-3.515)
Other	0.714	1.211	0.347	1	.56	2.041 (0.190-21.907)
Age (years)						
≤30	—	—	34.244	3	<.001	—
31-40	0.567	0.330	2.956	1	.09	1.763 (0.924-3.366)
41-50	–0.310	0.313	0.980	1	.32	0.733 (0.397-1.355)
>50	–0.845	0.309	7.469	1	.01	0.429 (0.234-0.787)
Ethnicity						
Caucasian	—	—	0.548	3	.91	—
Indigenous	0.358	.644	0.309	1	.58	1.430 (0.405-5.052)
Asian	–0.160	.368	0.189	1	.66	0.852 (0.414-1.753)
Other	–0.070	.337	0.043	1	.84	0.933 (0.482-1.805)
Education						
Less than high school diploma	—	—	1.442	3	.69	—
High school diploma	1.430	1.661	0.741	1	.39	4.177 (0.161-108.342)
Postsecondary education	0.751	1.282	0.343	1	.56	2.118 (0.172-26.144)
Other education	–.043	1.597	0.001	1	.98	0.958 (0.042-21.934)
Relationship status						
Married/cohabiting/partnered	—	—	1.239	4	.87	—
Separated/divorced	–.090	0.300	0.089	1	.77	0.914 (0.508-1.646)
Widowed	–.362	0.885	0.167	1	.68	0.696 (0.123-3.948)
Single	.253	0.297	0.730	1	.39	1.288 (0.720-2.304)
Other	–.444	1.272	0.122	1	.73	0.641 (0.053-7.758)
Housing status						
Own home	—	—	2.175	3	.54	—
Living with family	.743	0.656	1.285	1	.26	2.103 (0.582-7.599)
Renting	.296	0.265	1.244	1	.27	1.344 (0.799-2.261)
Other	20.278	15,888.453	0	1	.99	640,763,554.13 (0 to upper level)
Self-isolate/quarantine						
No	—	—	—	—	—	—
Yes	–.108	0.204	0.281	1	.59	0.898 (0.602-1.339)
HCW type						
Physician	—	—	2.874	2	.24	—
Nurses	.573	0.339	2.864	1	.09	1.774 (0.913-3.446)
Other	.474	0.321	2.180	1	.14	1.606 (0.856-3.012)
Constant	–.200	1.347	.022	1	.88	0.819 (—)

[a]HCW: health care worker.

[b]Significant P values are italicized.

[c]OR: odds ratio.

XSL•FO

RenderX

HCWs who were >50 years old were less likely to report moderate-to-high stress during the COVID-19 pandemic compared to respondents who were ≤30 years old, when all other variables in the model were controlled (OR 0.43, 95% CI 0.23-0.79). Although the gender variable did not significantly contribute to the model, females were almost 2 times more likely to report moderate-to-high stress during the COVID-19 pandemic compared to males, when controlling for other variables (OR 2.0, 95% CI 1.12-3.52).

The data in Table 3 indicate that for likely GAD, the 8-predictor model was statistically significant (χ^2_{21}=70.82, P<.001),

explaining between 7.3% (Cox and Snell R^2) and 9.9% (Nagelkerke R^2) of the variance and correctly classified 64.9% of all cases. The type of HCWs did not significantly contribute to likely GAD among all the HCWs. Likewise, age categories made a unique statistical contribution (Wald=41.85, P<.001) to the probability that a respondent presented with likely GAD, after controlling for all other factors in this model. HCWs who were 41-50 years old and those who were >50 years old had a lower probability of reporting likely GAD symptoms during the COVID-19 pandemic compared to those who were ≤30 years old (OR 0.41, 95% CI 0.26-0.66 vs OR 0.33, 95% CI 0.2-0.53).

Table 3. Logistic regression predicting likelihood for the respondents among HCWs[a] to present with likely GAD[b].

Variables	B	SE	Wald	df	P value[c]	OR[d] (95% CI)
Gender						
Male	__[e]	4.250	2	2	.12	—
Female	0.468	2.497	1	1	.11	1.597 (0.894-2.852)
Other	1.657	2.898	1	1	.09	5.243 (0.778-35.319)
Age (years)						
≤30	—	—	41.852	3	*<.001*	—
31-40	–0.045	0.224	0.040	1	.84	0.956 (0.616-1.483)
41-50	–0.887	0.243	13.344	1	*<.001*	0.412 (0.256-0.663)
>50	–1.121	0.250	20.131	1	*<.001*	0.326 (0.200-0.532)
Ethnicity						
Caucasian	—	—	2.910	3	.41	—
Indigenous	0.453	0.421	1.161	1	.28	1.574 (0.690-3.591)
Asian	–0.247	0.312	0.625	1	.43	0.781 (0.424-1.441)
Other	0.288	0.291	0.984	1	.32	1.334 (0.755-2.359)
Education						
Less than high school diploma	—	—	2.004	3	.57	—
High school diploma	–0.585	1.444	0.164	1	.69	0.557 (0.033--9.442)
Postsecondary education	–1.194	1.285	0.864	1	.35	0.303 (0.024-3.760)
Other education	–1.790	1.669	1.151	1	.28	0.167 (0.006-4.394)
Relationship status						
Married/cohabiting/partnered	—	—	2.865	4	.58	—
Separated/divorced	0.148	0.274	0.293	1	.59	1.160 (0.678-1.984)
Widowed	–0.673	1.110	0.367	1	.55	0.510 (0.058-4.498)
Single	0.293	0.220	1.768	1	.18	1.340 (0.870-2.064)
Other	–0.691	1.072	0.416	1	.52	0.501 (0.061-4.095)
Housing status						
Own home	—	—	3.756	3	.29	—
Living with family	0.027	0.412	0.004	1	.95	1.028 (0.458-2.305)
Renting	0.028	0.196	0.020	1	.89	1.028 (0.699-1.510)
Other	2.323	1.200	3.748	1	.05	10.210 (0.972-107.273)
Self-isolate/quarantine						
No	—	—	—	—	—	—
Yes	–0.236	0.172	1.879	1	.17	0.790 (0.563-1.107)
HCW type						
Physician	—	—	0.400	2	.82	—
Nurses	–0.113	0.330	0.117	1	.73	0.893 (0.467-1.706)
Other	–0.019	0.318	0.003	1	.95	0.981 (0.526-1.831)
Constant	0.848	1.351	0.394	1	.53	2.334 (—)

[a]HCW: health care worker.

[b]GAD: generalized anxiety disorder.

[c]Significant *P* values are italicized.

[d]OR: odds ratio.

[e]—: Not applicable.

The data displayed in Table 4 indicate that for likely MDD, the model containing the 8 predictors was statistically significant (χ^2_{21}=69.14, P<.001). The model explained between 7.0% (Cox and Snell R^2) and 9.8% (Nagelkerke R^2) of the variance and correctly classified 68% of all cases. After controlling for all other factors in the model, the type of HCWs made a unique statistical contribution (Wald=6.1, P=.05) to the likelihood that a respondent presented with moderate-to-high MDD. Nurses and other HCWs exhibited a 2 times greater probability of presenting with likely MDD during the pandemic compared to physicians (OR 2.32, 95% CI 1.06-5.10 vs OR 2.60, 95% CI 1.20-5.58). The variable of age categories made a unique statistical contribution (Wald=24.54, P<.001) to the probability that a respondent presented with likely MDD. Like the previous (GAD) model, HCWs who were 41-50 and >50 years old had a lower OR to report likely GAD symptoms during the COVID-19 pandemic compared to those ≤30 years old (OR 0.55, 95% CI 0.34-0.90 vs OR 0.39, 95% CI 0.24-0.65). Although the relationship status variable did not contribute significantly to the model, separated/divorced and single respondents reported a higher likelihood of MDD compared to married/cohabiting/partnered respondents when controlling for other variables (OR 1.80, 95% CI 1.04-3.02 vs OR 1.60, 95% CI 1.07-2.52).

From the 3 models collectively, there was a trend for age, whereby HCWs who were below 40 years old appeared more likely to exhibit psychological clinical symptoms compared to older participants.

Table 4. Logistic regression predicting likelihood for the respondents among HCWs[a] to present with likely MDDb.

Variables	B	SE	Wald	df	P value[c]	OR[d] (95% CI)
Gender						
Male	__e	—	0.619	2	.73	—
Female	0.031	0.294	0.011	1	.92	1.032
Other	0.714	0.918	0.606	1	.44	2.042
Age (years)						
≤30	—	—	24.539	3	*<.001*	—
31-40	–0.025	0.227	0.012	1	.91	0.975
41-50	–0.590	0.247	5.703	1	*.02*	0.554
>50	–0.934	0.258	13.111	1	*<.001*	0.393
Ethnicity						
Caucasian	—	—	1.282	3	.73	—
Indigenous	0.046	0.427	0.011	1	.92	1.047
Asian	–0.075	0.321	0.054	1	.82	0.928
Other	0.319	0.295	1.168	1	.28	1.376
Education						
Less than high school diploma	—	—	3.471	3	.32	—
High school diploma	–0.677	1.406	0.232	1	.63	0.508
Postsecondary education	–1.500	1.272	1.390	1	.24	0.223
Other education	–0.950	1.640	0.336	1	.56	0.387
Relationship status						
Married/cohabiting/partnered	—	—	8.863	4	.07	—
Separated/divorced	0.574	0.271	4.493	1	*.03*	1.776
Widowed	–0.273	1.109	0.060	1	.81	0.761
Single	0.495	0.219	5.107	1	*.02*	1.640
Other	–0.537	1.055	0.259	1	.61	0.584
Housing status						
Own home	—	—	4.775	3	.19	—
Living with family	.356	.413	0.744	1	.39	1.428
Renting	.225	.195	1.326	1	.25	1.252
Other	2.195	1.201	3.340	1	.07	8.980
Self-isolate/quarantine						
No	—	—	—	—	—	—
Yes	–0.114	0.177	0.418	1	0.52	0.892
HCW type						
Physician	—	—	6.076	2	*.05*	—
Nurses	0.843	0.402	4.396	1	*.04*	2.322
Other	0.952	0.391	5.932	1	*.02*	2.592
Constant	0.095	1.358	0.005	1	.94	1.099 (—)

[a]HCW: health care worker.

[b]MDD: major depressive disorder.

[c]Significant P values are italicized.

[d]OR: odds ratio.

[e]—: Not applicable.

Discussion

Principal Findings

Using self-reported data from the Text4Hope service, this study illustrates the different mental health impacts of the COVID-19 pandemic on HCWs, including physicians, nurses, and other HCWs. The study suggests that more than 4 in 5 HCWs expressed a likelihood of reporting moderate-to-high perceived stress rates, while around a third expressed a likelihood of reporting moderate-to-severe anxiety and depressive symptoms during the COVID-19 pandemic. In Muller et al's [18] systematic review, it was determined that the prevalence of mental health distress ranged from 7% to 97%, with a median of 37%. In our study, the prevalence of GAD and depression in HCWs was 38.6% and 32.7%, respectively, which was comparable to the medians reported in the aforementioned 2 systematic reviews [18,19].

In a survey of health care providers in Wuhan, the number of frontline health care providers with depressive symptoms was estimated to be 50.4%, with correlates of nursing profession and female gender [20]. This higher figure compared with the prevalence of depression in our study may be explained by the demographics of the population that study focused on, as most participants were female, nurses, of lower ages, and with a junior technical title that could be considered as elevating COVID-19 exposure risk factors for the development of depressive symptoms [20,27].

Physicians had the lowest prevalence of GAD and depression in this study (20/63 [36.4%] and 9/63 [16.1%], respectively), followed by nurses (116/355 [37.4%] and 99/355 [31.3%], respectively) and was highest in other HCW groups (233/678 [39.4%] and 209/678 [35.1%], respectively). Pappa et al [19] noted similar trends in the prevalence of depression and anxiety between physicians and nurses in their systematic review. Physicians had a lower prevalence of depression (25.4%) when compared with nurses (30.3%), and this was the case for GAD that was found in 21.7% of physicians, while 25.8% of nurses were affected [19].

Compared to the frontline physicians, first responder nurses are more likely to develop behavioral disengagement during similar epidemics, a result highly associated with aggravating levels of depression [28,29]. This may be due to stress and the fear of contracting the infection or spreading it to their families, along with the perceived stigma and sense of uncertainty [20]. Consistent with our research, where nurses were found to be more likely to report moderate-to-severe likely stress symptoms compared to the physicians, in a study carried out in Wuhan, China, nurses were also reported to experience higher levels of different psychological impacts, such as distress, anxiety, depression, and insomnia, when compared to their physician counterparts during the COVID-19 pandemic [20].

Females comprised most of our survey respondents (1006/1096, 92%) and showed a higher likelihood to report stress, GAD, and depressive symptoms compared to their male counterparts. However, this was statistically significant only for likely stress symptoms. This was consistent with other studies. For example, after pooling the prevalence from 6 studies that reported complete data for anxiety symptoms based on gender during the COVID-19 pandemic, Pappa et al [19] reported a lower prevalence of anxiety in males (20.9%) compared to females (29.1%). Conversely, they noted that the prevalence of depression was higher in males (26.9%) than in females (20.3%) [19]. Several studies have demonstrated that being a woman, getting exposed to patients with SARS-CoV-2 infection, and worrying about being infected are the most common risk factors associated with increased mental health problems in HCWs during the pandemic [18].

Younger HCWs were generally at a higher risk of developing likely stress, anxiety, and depression compared to other groups. This finding is aligned with other previous findings that consistently report a higher risk among younger ages during the pandemic [27,30,31]. This may be attributed to the longer duration the younger generation usually spends focusing upon the data of the pandemic or to the lack of experiencing similar stressful situations in their few years of experience compared to older groups, who possibly have passed by similar experiences [27,30].

In our study, the lack of a confiding relationship (separated/divorced and single) appeared to be a risk factor for HCWs to self-report likely depression. It is likely that HCWs who are in confiding relationships have better social supports that those who are single, divorced, or widowed. This finding appears clearly in the literature, where social, family, and friend support is found to be the most commonly reported protective factor associated with a reduction in mental health problems among HCWs [18,32].

Not surprisingly, the younger age group of HCWs expressed a higher likelihood to show likely stress, GAD, and MDD symptoms compared to the other older groups. Additionally, the likelihood to report the 3 clinical conditions of focus in our study seemed to decrease with age, where HCWs who were ≤30 years old were at a higher risk of developing likely stress, GAD, and MDD during the pandemic compared to those who were 41-50 and >50 years old (OR 1.82-3.03). This aligns with previous research that reported the same finding among all subscribers of Text4Hope [27,33]. Relatively little experience and the lack of exposure to similar epidemics in the younger age group could predispose younger people to a greater likelihood of developing maladjustment and mental distress in this context [27,34].

Other researchers are examining COVID-19-related measurements among HCWs across Canada, including exposure to the virus and other mental health parameters [13]. Preliminary results from the other study suggests that there are high COVID-19 infection rates among HCWs and that physicians are more likely than other HCWs to develop mental health symptoms. These results differ from the findings of our study, which suggest that physicians are less likely to develop mental health symptoms during the COVID-19 pandemic compared to nurses and other HCWs. Observed differences between each

study's results could be attributed to differences in data collection periods (eg, early pandemic, during a wave), differences in COVID-19 positivity rates and R values, or even differences in the level of resource strain (eg, bed capacity, worker shortages) on the health care system. Although our data were collected during the early phase of the pandemic, when the health system was not strained, data from an ongoing Canadian study were collected at a later stage of the pandemic, when infection rates were higher and health care systems across Canada were under greater strain.

During the COVID-19 pandemic, some digital initiatives took place to mitigate psychological distress among health care professionals; these included a digital support package on psychological well-being provided for HCWs in the United Kingdom. The package was easily delivered and reached the target group, with a high rate of usage and accepted cost [7]. In a recent review of the literature demonstrating the mental health problems that HCWs are facing during the COVID-19 pandemic, the authors concluded that health authorities need to build multidisciplinary mental health teams in order to mitigate mental health and psychological consequences of the pandemic on both patients and HCWs. It was suggested that electronic media through web apps could be used for this purpose [6]. In Alberta, Canada, the AHS launched a supportive text message program (Text4Hope). The service aims to support the mental health of Albertans during the COVID-19 pandemic and is presented as a feasible and reliable medium for wide-scale data collection for epidemiological research, as illustrated by this report.

Limitations

This study had several limitations. The data were obtained via an online request sent to all subscribers of the Text4Hope service, and there was a risk of selection bias, where the HCWs who responded may have been more interested in the service and not necessarily representative of all HCWs in Alberta. Females also showed overrepresentation in our cohort, which may limit the generalization of the provided results. Additionally, as self-report scales were used in this study and participants were not clinically assessed or confirmed, the results need to be interpreted carefully.

Finally, this was a cross-sectional study with no established baseline prevalence or a control group to compare with. It cannot draw any inferences about the impact of the pandemic on HCWs' mental health; thus, the presented findings may need further similar studies across different timepoints into the pandemic to validate our results.

Conclusion

This cross-sectional study explored the impact of the COVID-19 pandemic on the mental health and well-being of HCWs in Alberta. Overall, we highlighted the prevalence of the psychological distress symptoms in the early phase of the COVID-19 pandemic. Nurses were observed to be at high risk for developing stress, depression, and anxiety during the pandemic. Tracking such symptoms through digitally supported means is highly emphasized, particularly during pandemics where physical contact is not a viable option. Usually, digitally provided programs may yield high fidelity in relation to their acceptability and engagement [35]. Such services are impactful and cost-effective and fulfil essential social distancing requirements with remote delivery.

In the light of time progression and the development of the knowledge around the pandemic and the available vaccination, we aim to examine the changes in these symptoms and their timely progress among the health care providers after 1 year of the initiation of the Text4Hope service. From the literature, positive impacts of similar interventions were reported, in terms of reducing depressive symptoms and increasing the abstinence duration in alcohol use disorder, after 3 months of receiving daily supportive text messages in the community settings [36,37]. The availability of the texting service on most cell phones, and the lack of required software or app downloads to function, rendered text messaging services significantly advantageous over similar technologies (eg, email or messaging apps) [36].

References

1. Kang L, Ma S, Chen M, Yang J, Wang Y, Li R, et al. Impact on mental health and perceptions of psychological care among medical and nursing staff in Wuhan during the 2019 novel coronavirus disease outbreak: a cross-sectional study. Brain Behav Immun 2020 Jul;87:11-17 [FREE Full text] [doi: 10.1016/j.bbi.2020.03.028] [Medline: 32240764]

2. WHO Director-General's Opening Remarks at the Media Briefing on COVID-19 - 1 February 2022. Geneva, Switzerland: WHO URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2022-02-07]

3. Health Care Glossary. URL: https://www.albertahealthservices.ca/about/Page12677.aspx [accessed 2022-02-07]

4. Canada's Health Care Providers. URL: https://secure.cihi.ca/free_products/hctenglish.pdf [accessed 2022-02-07]

5. COVID-19 Cases and Deaths among Health Care Workers in Canada, 2020. URL: https://www.cihi.ca/en/covid-19-cases-and-deaths-among-health-care-workers-in-canada [accessed 2022-02-07]

6. Spoorthy MS, Pratapa SK, Mahant S. Mental health problems faced by healthcare workers due to the COVID-19 pandemic: a review. Asian J Psychiatr 2020 Jun;51:102119 [FREE Full text] [doi: 10.1016/j.ajp.2020.102119] [Medline: 32339895]

7. Blake H, Bermingham F, Johnson G, Tabner A. Mitigating the psychological impact of COVID-19 on healthcare workers: a digital learning package. Int J Environ Res Public Health 2020 Apr 26;17(9):2997 [FREE Full text] [doi: 10.3390/ijerph17092997] [Medline: 32357424]

8. Wu K, Wei X. Analysis of psychological and sleep status and exercise rehabilitation of front-line clinical staff in the fight against COVID-19 in China. Med Sci Monit Basic Res 2020 May 11;26:e924085. [doi: 10.12659/msmbr.924085]

9. McAlonan GM, Lee AM, Cheung V, Cheung C, Tsang KW, Sham PC, et al. Immediate and sustained psychological impact of an emerging infectious disease outbreak on health care workers. Can J Psychiatry 2007 Apr 01;52(4):241-247. [doi: 10.1177/070674370705200406] [Medline: 17500305]

10. Di Tella M, Romeo A, Benfante A, Castelli L. Mental health of healthcare workers during the COVID-19 pandemic in Italy. J Eval Clin Pract 2020 Dec 25;26(6):1583-1587. [doi: 10.1111/jep.13444] [Medline: 32710481]

11. Maunder R, Hunter J, Vincent L, Bennett J, Peladeau N, Leszcz M, et al. The immediate psychological and occupational impact of the 2003 SARS outbreak in a teaching hospital. CMAJ 2003 May 13;168(10):1245-1251 [FREE Full text] [Medline: 12743065]

12. Ornell F, Halpern SC, Kessler FHP, Narvaez JCDM. The impact of the COVID-19 pandemic on the mental health of healthcare professionals. Cad Saúde Pública 2020;36(4):e00063520. [doi: 10.1590/0102-311x00063520]

13. University of Alberta Study Examines COVID-19 Toll on Health-care Workers. CBC News. URL: https://www.cbc.ca/news/canada/edmonton/u-of-a-study-mental-health-care-covid-1.5818020#:~:text=An%20ongoing%20study%20at%20the,British%20Columbia%2C%20Ontario%20and%20Quebec [accessed 2022-02-07]

14. Patel R, Bachu R, Adikey A, Malik M, Shah M. Factors related to physician burnout and its consequences: a review. Behav Sci (Basel) 2018 Oct 25;8(11):98 [FREE Full text] [doi: 10.3390/bs8110098] [Medline: 30366419]

15. Greenberg N, Docherty M, Gnanapragasam S, Wessely S. Managing mental health challenges faced by healthcare workers during COVID-19 pandemic. BMJ 2020 Mar 26;368:m1211. [doi: 10.1136/bmj.m1211] [Medline: 32217624]

16. Alshekaili M, Hassan W, Al Said N, Al Sulaimani F, Jayapal SK, Al-Mawali A, et al. Factors associated with mental health outcomes across healthcare settings in Oman during COVID-19: frontline versus non-frontline healthcare workers. BMJ Open 2020 Oct 10;10(10):e042030 [FREE Full text] [doi: 10.1136/bmjopen-2020-042030] [Medline: 33040019]

17. Chew NW, Lee GK, Tan BY, Jing M, Goh Y, Ngiam NJ, et al. A multinational, multicentre study on the psychological outcomes and associated physical symptoms amongst healthcare workers during COVID-19 outbreak. Brain Behav Immun 2020 Aug;88:559-565 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.049] [Medline: 32330593]

18. Muller AE, Hafstad EV, Himmels JPW, Smedslund G, Flottorp S, Stensland, et al. The mental health impact of the COVID-19 pandemic on healthcare workers, and interventions to help them: a rapid systematic review. Psychiatry Res 2020 Nov;293:113441 [FREE Full text] [doi: 10.1016/j.psychres.2020.113441] [Medline: 32898840]

19. Pappa S, Ntella V, Giannakas T, Giannakoulis VG, Papoutsi E, Katsaounou P. Prevalence of depression, anxiety, and insomnia among healthcare workers during the COVID-19 pandemic: a systematic review and meta-analysis. Brain Behav Immun 2020 Aug;88:901-907 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.026] [Medline: 32437915]

20. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors associated with mental health outcomes among health care workers exposed to coronavirus disease 2019. JAMA Netw Open 2020 Mar 02;3(3):e203976 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3976] [Medline: 32202646]

21. Agyapong VIO, Hrabok M, Vuong W, Gusnowski A, Shalaby R, Mrklas K, et al. Closing the psychological treatment gap during the COVID-19 pandemic with a supportive text messaging program: protocol for implementation and evaluation. JMIR Res Protoc 2020 Jun 22;9(6):e19292. [doi: 10.2196/19292]

22. Agyapong VIO. Coronavirus disease 2019 pandemic: health system and community response to a text message (Text4Hope) program supporting mental health in Alberta. Disaster Med Public Health Prep 2020 Oct 22;14(5):e5-e6 [FREE Full text] [doi: 10.1017/dmp.2020.114] [Medline: 32317038]

23. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385. [doi: 10.2307/2136404]

XSL•FO

RenderX

24. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

25. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

26. IBM Corp. Released 2011. IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp; 2011:2011.

27. Nwachukwu I, Nkire N, Shalaby R, Hrabok M, Vuong W, Gusnowski A, et al. COVID-19 pandemic: age-related differences in measures of stress, anxiety and depression in Canada. Int J Environ Res Public Health 2020 Sep 01;17(17):6366 [FREE Full text] [doi: 10.3390/ijerph17176366] [Medline: 32882922]

28. Burker EJ, Evon D, Loiselle MM, Finkel J, Mill M. Planning helps, behavioral disengagement does not: coping and depression in the spouses of heart transplant candidates. Clin Transplant 2005 Oct;19(5):653-658. [doi: 10.1111/j.1399-0012.2005.00390.x] [Medline: 16146558]

29. Wong TW, Yau JK, Chan CL, Kwong RS, Ho SM, Lau CC, et al. The psychological impact of severe acute respiratory syndrome outbreak on healthcare workers in emergency departments and how they cope. Eur J Emerg Med 2005 Feb;12(1):13-18. [doi: 10.1097/00063110-200502000-00005] [Medline: 15674079]

30. Huang Y, Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: a web-based cross-sectional survey. Psychiatry Res 2020 Jun;288:112954 [FREE Full text] [doi: 10.1016/j.psychres.2020.112954] [Medline: 32325383]

31. Lei L, Huang X, Zhang S, Yang J, Yang L, Xu M. Comparison of prevalence and associated factors of anxiety and depression among people affected by versus people unaffected by quarantine during the COVID-19 epidemic in Southwestern China. Med Sci Monit 2020 Apr 20;26:e924609. [doi: 10.12659/msm.924609]

32. Liu X, Shao L, Zhang R, Wei Y, Li J, Wang C, et al. Perceived social support and its impact on psychological status and quality of life of medical staffs after outbreak of SARS-CoV-2 pneumonia: a cross-sectional study. SSRN J 2020:21 pages. [doi: 10.2139/ssrn.3541127]

33. Hrabok M, Nwachukwu I, Gusnowski A, Shalaby R, Vuong W, Surood S, et al. Mental health outreach via supportive text messages during the COVID-19 pandemic: one-week prevalence and correlates of anxiety symptoms. Can J Psychiatry 2021 Jan 02;66(1):59-61 [FREE Full text] [doi: 10.1177/0706743720969384] [Medline: 33131318]

34. An Y, Yang Y, Wang A, Li Y, Zhang Q, Cheung T, et al. Prevalence of depression and its impact on quality of life among frontline nurses in emergency departments during the COVID-19 outbreak. J Affect Disord 2020 Nov 01;276:312-315 [FREE Full text] [doi: 10.1016/j.jad.2020.06.047] [Medline: 32871661]

35. Bohlken J, Schömig F, Lemke MR, Pumberger M, Riedel-Heller SG. COVID-19 pandemic: stress experience of healthcare workers - a short current review. Psychiatr Prax 2020 May;47(4):190-197 [FREE Full text] [doi: 10.1055/a-1159-5551] [Medline: 32340048]

36. Agyapong VIO, Mrklas K, Juhás M, Omeje J, Ohinmaa A, Dursun SM, et al. Cross-sectional survey evaluating Text4Mood: mobile health program to reduce psychological treatment gap in mental healthcare in Alberta through daily supportive text messages. BMC Psychiatry 2016 Nov 08;16(1):378 [FREE Full text] [doi: 10.1186/s12888-016-1104-2] [Medline: 27821096]

37. Agyapong V, McLoughlin D, Farren C. 581 – 6-month outcomes of a single blind randomised trial of supportive text messaging for depression and comorbid alcohol use disorder. Eur Psychiatry 2013 Jan;28:1. [doi: 10.1016/s0924-9338(13)75859-0]

Abbreviations

AHS: Alberta Health Services
GAD: generalized anxiety disorder
GAD-7: Generalized Anxiety Disorder 7-item
HCW: health care worker
MDD: major depressive disorder
OR: odds ratio
PHQ-9: Patient Health Questionnaire-9
PSS-10: Perceived Stress Scale

XSL•FO
RenderX

XSL•FO
RenderX

Original Paper

System-Level Factors Associated With Telephone and Video Visit Use: Survey of Safety-Net Clinicians During the Early Phase of the COVID-19 Pandemic

Anjana E Sharma[1,2], MAS, MD; Elaine C Khoong[2,3], MS, MD; Maribel Sierra[2,3], MPH; Natalie A Rivadeneira[2,3], MPH; Malini A Nijagal[4], MPH, MD; George Su[5], MD; Courtney R Lyles[2,3], PhD; Triveni DeFries[3], MPH, MD; Delphine S Tuot[2,6], MAS, MDCM; Urmimala Sarkar[2,3], MPH, MD

[1]Department of Family & Community Medicine, University of California San Francisco, San Francisco, CA, United States

[2]Center for Vulnerable Populations, Zuckerberg San Francisco General Hospital, University of California San Francisco, San Francisco, CA, United States

[3]Division of General Internal Medicine, Zuckerberg San Francisco General Hospital, Department of Medicine, University of California San Francisco, San Francisco, CA, United States

[4]Department of Obstetrics and Gynecology, University of California San Francisco, San Francisco, CA, United States

[5]Division of Pulmonary, Critical Care, and Sleep, Zuckerberg San Francisco General Hospital, Department of Medicine, University of California San Francisco, San Francisco, CA, United States

[6]Department of Nephrology, University of California San Francisco, San Francisco, CA, United States

Corresponding Author:
Anjana E Sharma, MAS, MD
Department of Family & Community Medicine
University of California San Francisco
995 Potrero Ave, Ward 83
San Francisco, CA, 94110
United States
Phone: 1 4155708619
Email: anjana.sharma@ucsf.edu

Abstract

Background: The COVID-19 pandemic prompted safety-net health care systems to rapidly implement telemedicine services with little prior experience, causing disparities in access to virtual visits. While much attention has been given to patient barriers, less is known regarding system-level factors influencing telephone versus video-visit adoption. As telemedicine remains a preferred service for patients and providers, and reimbursement parity will not continue for audio visits, health systems must evaluate how to support higher-quality video visit access.

Objective: This study aimed to assess health system–level factors and their impact on telephone and video visit adoption to inform sustainability of telemedicine for ambulatory safety-net sites.

Methods: We conducted a cross-sectional survey among ambulatory care clinicians at a hospital-linked ambulatory clinic network serving a diverse, publicly insured patient population between May 28 and July 14, 2020. We conducted bivariate analyses assessing health care system–level factors associated with (1) high telephone adoption (4 or more visits on average per session); and (2) video visit adoption (at least 1 video visit on average per session).

Results: We collected 311 responses from 643 eligible clinicians, yielding a response rate of 48.4%. Clinician respondents (N=311) included 34.7% (n=108) primary or urgent care, 35.1% (n=109) medical, and 7.4% (n=23) surgical specialties. Our sample included 178 (57.2%) high telephone adopters and 81 (26.05%) video adopters. Among high telephone adopters, 72.2% utilized personal devices for telemedicine (vs 59.0% of low telephone adopters, *P*=.04). Video nonadopters requested more training in technical aspects than adopters (49.6% vs 27.2%, *P*<.001). Primary or urgent care had the highest proportion of high telephone adoption (84.3%, compared to 50.4% of medical and 37.5% of surgical specialties, *P*<.001). Medical specialties had the highest proportion of video adoption (39.1%, compared to 14.8% of primary care and 12.5% of surgical specialties, *P*<.001).

Conclusions: Personal device access and department specialty were major factors associated with high telephone and video visit adoption among safety-net clinicians. Desire for training was associated with lower video visit use. Secure device access, clinician technical trainings, and department-wide assessments are priorities for safety-net systems implementing telemedicine.

Introduction

The COVID-19 pandemic catalyzed a dramatic increase in telemedicine care, prompted by the need for physical distancing and Medicare and Medicaid reimbursement changes enabling parity in coverage for telemedicine [1-3]. Many safety-net and public health systems serving primarily publicly insured, low-income populations [4,5], implemented ambulatory telemedicine with little prior preparation or experience [6,7]. Most safety-net sites provided mainly audio-only telephone visits, in contrast to non–safety-net sites that had prior video visit infrastructure [8]. Video visits have higher patient satisfaction than telephone consultations [9], and policy briefs [10] predict that telephone-only visits will not be reimbursed at parity with video after the pandemic abates [11,12]. As telemedicine becomes integrated into ambulatory care, safety-net health networks must explore factors affecting audio versus video telemedicine visits to maintain reimbursement parity with in-person visits.

Prior studies have found older adults, those insured by Medicaid, or low-income [13] and racial and ethnic minorities receive fewer video visits [14-16]. This has been hypothesized to be due to patient-facing barriers, such a lower digital literacy and lack of patient access to video-enabled smartphone and internet [17,18] and, to some degree, clinician-specific factors. However, a recent analysis of telephone and video visit variation found that practices (38%) and clinicians (26%) drove more of the variation in video visit use than patient-level factors (9%) [18]. System workflows are emphasized as influential for telemedicine implementation in rural sites [19]. Since much of the current literature has focused on patient-level telemedicine barriers to explain differences in telephone versus video visit uptake in the safety-net [9,18,20], we sought to assess other system-level factors influencing telephone and video visit implementation at a large urban public hospital during the first 6 months of the COVID-19 pandemic. Examination of system-level implementation factors during the early transitional phase of telemedicine can inform the adoption and sustainability of video visit use at safety-net sites [13].

Methods

Study Setting

Our study setting was a large, hospital-linked ambulatory clinic network serving an ethnically diverse, publicly insured patient population. This network had limited telemedicine visit capacity prior to the COVID-19 pandemic and began providing telemedicine care on March 3, 2020, with support and infrastructure primarily for telephone visits. Departments and clinicians could provide video visits on an ad hoc basis. All clinic and charting rooms had landlines and access to local area network–connected client computers without webcams. The network developed a standardized electronic health record (EHR) visit type and video visit workflow on June 15, 2020. Details of EHR implementation for telemedicine during the COVID-19 pandemic in this network are published elsewhere [21].

Ethical Considerations

This quality improvement study was exempt from institutional review board review.

Study Participants

We invited all 643 clinicians providing ambulatory telemedicine care to participate via email, with prompts from specialty-specific study champions. We aimed for a response rate greater than 40% given the challenges of surveying clinicians during the COVID-19 pandemic.

Study Instrument

We conducted a cross-sectional anonymous survey via the cloud-based platform Qualtrics. We developed our survey on the basis of key constructs from validated implementation science and telemedicine surveys such as the Telehealth Usability Questionnaire [22-25] and discussions with clinical leaders in the network. The full survey is provided in Multimedia Appendix 1. We conducted pretesting of the instrument at 2 clinical sites from April 20 to May 18, 2020. We distributed the finalized survey from May 28 to July 14, 2020. Reminders were sent by department-specific champions during June 2020.

Primary Outcome Variables: Phone and Video Visit Adoption

To assess telemedicine adoption, we asked respondents, "On average, how many telemedicine visits do you complete per half-day session? Think back to the last month of ambulatory care." Answer options were 0, 1-3, 4-6, 7-9, or ≥10. We dichotomized responses for our two distinct outcomes of interest: (1) high telephone visit adoption (defined as 4 visits per half day or more on average in the last month) and (2) video visit adoption (defined as at least 1 video visit on average per half day in the last month at the telemedicine clinic). We defined "high telephone adopter" as 4 visits or more because telemedicine was still a novel innovation for the delivery system at this time, with almost none prior to March 2020. In June 2020, ambulatory providers had 2-12 visits total per day across specialties (mean 5 visits per day); 4 visits comprise a majority of a day's visits and therefore fits our definition of substantial use. The half day distinction is because of the high proportion of part-time clinicians in this network. For video visits, there was extremely low uptake (<1%) of video visit use across the network [26]; therefore, even experience with 1 video visit per half day on average would signify adopters of video visits. We excluded residents and trainees as they rotated through multiple sites and had inconsistent experiences with telemedicine during this period.

Individual-Level Clinician Characteristics

We assessed demographic characteristics including age, gender, and clinical specialty. Respondents reported their age range (<30 years, 30-49 years, etc.), which we collapsed to a binary variable of <50 versus >50. We categorized gender as male and nonmale (combining female, nonbinary or nonconforming, and transgender entries to minimize identifiability without excluding sexual or gender minorities [27]). We grouped clinical specialty as "primary/urgent care," "medical specialty," or "surgical specialty" (see Table 1 and Multimedia Appendix 2 for all specialty categories included in this survey).

Table 1. Clinician characteristics and self-reported telemedicine use (N=311).

Characteristics	Participants, n (%)
Age (years)	
20-29	7 (2.25)
30-39	47 (15.1)
40-49	42 (13.5)
50-59	50 (16.1)
60-69	43 (13.8)
≥70	52 (16.7)
Missing or not disclosed	70 (22.5)
Gender	
Female	181 (58.2)
Male	51 (16.4)
Nonbinary	3 (0.9)
Transgender	0 (0)
Missing or not disclosed	76 (24.4)
Clinician role	
Faculty or attending physician	144 (46.3)
Nurse practitioner or physician assistant	51 (16.4)
Licensed counselor, social worker, or marriage family therapist	9 (2.9)
Other[a]	36 (11.5)
Missing or not disclosed[a]	71 (22.8)
Specialty	
Primary care and urgent care (total)	**108 (34.7)**
Family medicine	47 (15.1)
Internal medicine	32 (10.3)
Pediatrics	23 (7.4)
Other primary care[a]	6 (2.2)
Medical specialty (total)	**115 (37.0)**
Psychiatry	33 (10.6)
Obstetrics, gynecology, or midwifery	26 (8.4)
Oncology	9 (2.9)
Other medical specialty[a]	47 (15.1)
Surgical specialty (total)	**24 (7.7)**
Orthopedics	9 (2.9)
General surgery and trauma	5 (1.6)
Neurosurgery	2 (0.6)
Other surgical specialty[a]	8 (2.6)
Not disclosed	64 (20.6)
Telephone visits per half day	
0	38 (12.2)
1-3	96 (30.6)
4-6	100 (32.1)

Characteristics	Participants, n (%)
7-9	61 (19.6)
≥10	17 (5.5)
Video visits per half day	
0	230 (73.9)
1-3	71 (22.8)
4-6	7 (2.3)
7-9	0 (0.0)
≥10	3 (0.9)
Device used (smartphone, landline telephone, desktop, laptop, or tablet)	
Institutional only	77 (24.8)
At least one personal device	163 (52.4)
Missing	71 (22.8)

[a]See Multimedia Appendix 2 for all specialty categories included in this survey.

System-Level Variables

System-level factors included the following: perceived workload of telemedicine compared to in-person consultations (more, same, or less), estimated time spent helping patients navigating their telemedicine visit (dichotomized to <5 or ≥5 minutes), ease of interpreter use during telemedicine (Likert scale of 1-4 points, dichotomized to more or less difficult than in-person consultations), desire for additional telemedicine training (with respect to conducting technical aspects of visits, supporting patients with low digital literacy, gathering clinical information, developing an assessment or plan, or teaching trainees), and adequacy of audio or video quality of telemedicine encounters (yes or no). We asked what devices were used for telemedicine encounters; respondents could select smartphone, landline phone, desktop, laptop, or tablet device and specify if these were personal devices or institutional (work-provided) devices; we dichotomized this to "any personal device use" versus "only institutional device use."

Statistical Analysis

We conducted bivariate chi-square tests to assess if individual-level characteristics and system-level variables were associated with high telephone adoption, and, separately, video visit adoption. We conducted a sensitivity analysis including only respondents who had completed ≥50% of the survey for a "complete case" analysis. There were no major differences in including all responses versus "complete case" analysis; therefore, we report all responses available.

Statistical analysis was conducted in SAS software (version 9.4). This was a voluntary study encompassing the entire source population; no power calculation was attempted given the lack of prior reference data on telemedicine during a pandemic to extrapolate to the analysis.

Results

We collected 311 responses from 643 eligible clinicians, which resulted in a response rate of 48.4%. Full demographic characteristics of respondents are listed in Table 1. Primary and urgent care clinicians comprised 34.7%, medical specialists comprised 37.0%, and surgical specialists comprised 7.7% of the participant pool. Of clinicians, 57.2% (178/311) reported ≥4 telephone visits per half day on average. There were 81 of 311 (26.1%) clinicians who reported at least 1 video visit per half day on average. Most respondents (163, 52.4%) used ≥1 personal device (either smartphone, landline phone, laptop, desktop, or tablet device) for telemedicine encounters; 77 (24.8%) physicians used exclusively work-provided, institutional devices.

On bivariate analysis, we found that personal device use was associated with high telephone adoption (Figure 1A; Table 2). Among high telephone adopters, 72.2% (117/163) utilized at least one personal device for telemedicine encounters (P=.04). Personal device use was higher among video visit nonadopters than among video visit adopters, but this was not significant (71.2%, 126/177 vs 58.7%, 37/63, P=.07).

Many video adopters as well as nonadopters expressed interest in many training domains, especially supporting patients with low technical literacy, technical aspects of the visit, and teaching trainees telemedicine (Figure 2). Desire for training in conducting technical aspects of a telemedicine visit was statistically significantly higher for video nonadopters (49.6%) than for video adopters (27.2%, P<.001).

We found that clinical specialty was significantly associated with high telephone adoption and video adoption (Figure 1B). Primary or urgent care clinicians had the highest proportion of high telephone adoption (84.3%, 91/108), compared to 50.4% (58/115) of medical and 37.5% (9/24) of surgical specialties (P<.001). Medical specialties had the highest proportion of video adoption (39.1%, 45/115), compared to 14.8% (16/108) of primary care and 12.5% (3/24) of surgical physicians (P<.001).

We did not find associations between telemedicine adoption and workflow-specific challenges (Table 2). The perceived workload of telemedicine relative to in-person visits, the

estimated average time spent helping set up a patient for a telemedicine visit, and the perceived difficulty of working with an interpreter in a telemedicine encounter were not significantly different between telephone high-adopters or video adopters versus low or nonadopters. Few clinicians reported inadequate audio quality during telephone visits (16/166, 9.9% for high-adopters and 9/91, 12.0% for low adopters). Of video visit adopters, 8 of 70 (13.6%) found audiovisual quality inadequate.

Figure 1. Association between clinician and system-level factors with self-reported high telephone use and any video use, comparing adopters to nonregular/nonadopters. Experience with 4 completed telephone visits in a half day session implies a high telephone adopter, and <4 implies a low telephone visit adopter. 1 video visit per half day on average signifies a video visit adopter. (A) Using at least one personal device is associated with being a high telephone adopter (72.2%, 117/162), compared to 59.0% (46/78) among low telephone adopters (χ^2_1= 4.24, P=.04). Personal device use was higher among video nonadopters than for video visit adopters, but this was not significant (58.7%, 37/63 vs 71.2%, 126/177; χ^2_1=3.31; P=.07). (B) Primary or urgent care specialty had the greatest high telephone adoption (84.3%, 91/108) compared to medical (50.4%, 58/115) and surgical (37.5%, 9/24) specialties (χ^2_1=35.7, P<.001). Medical specialties had the highest proportion of video adoption (39.1%, 45/115) compared to primary care (14.8%, 16/108) and surgical (12.5%, 3/24) (χ^2_1=19.64, P<.001).
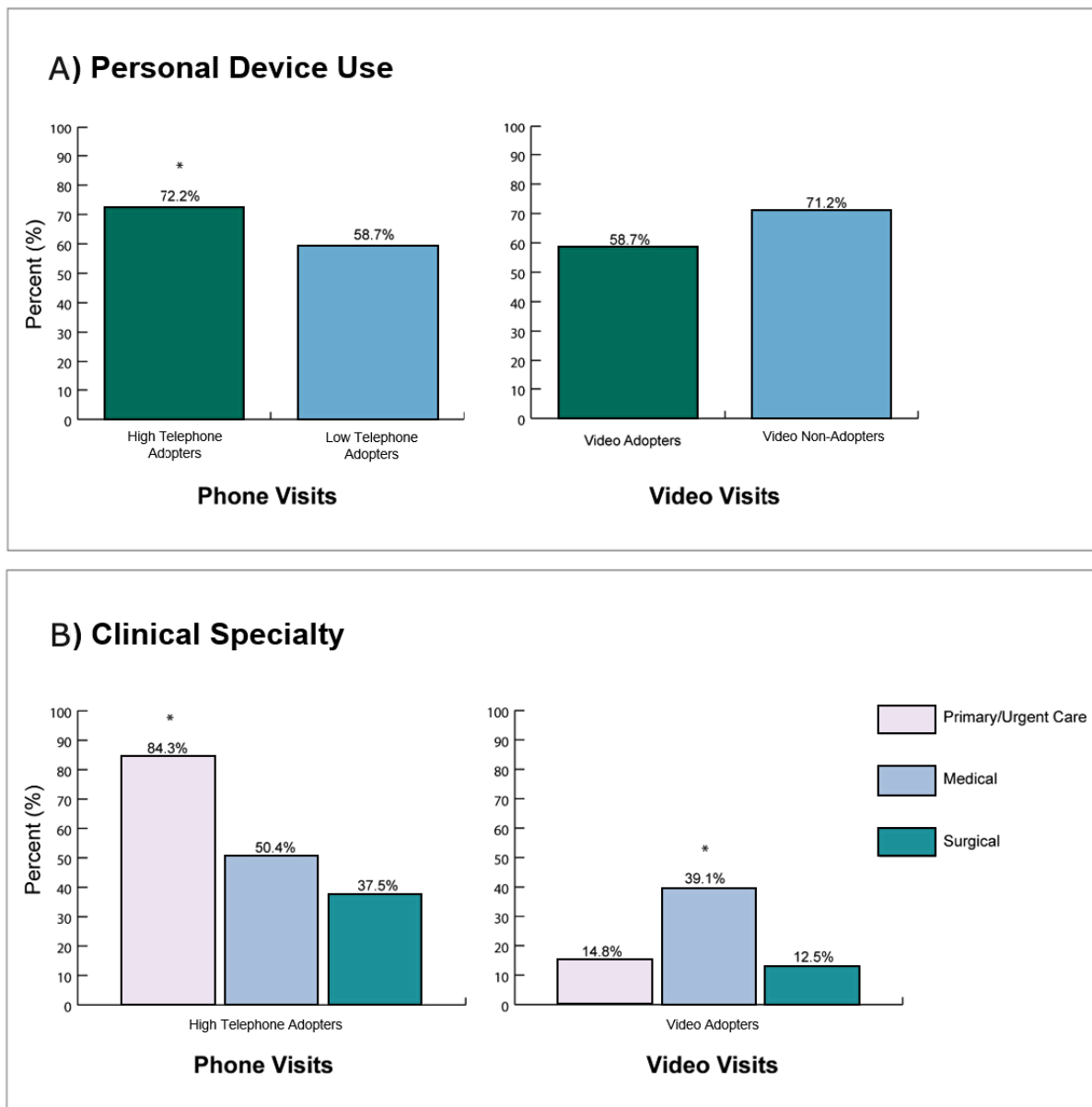
Table 2. Association of system-level factors with self-reported high telephone use and any video use, the latter 2 being independent outcomes and calculated separately.

System-level factors (N=311)	Telephone visits per half day				Video visits per half day			
	≥4 visits (n=178), n/n (%)	≤3 visits (n=133), n/n (%)	Chi-square (df)	P value[a]	≥1 video visit (n=81), n/n (%)	No video visit (n=230), n/n (%)	Chi-square (df)	P value[a]
Perceived workload (compared to in-person visits)			2.78 (1)	.10			1.32 (1)	.25
Less or same workload	143/166 (86.1)	71/91 (78.0)			42/70 (60.0)	20/41 (48.8)		
More workload	23/166 (13.9)	20/91 (22.0)			28/70 (40.0)	21/41 (51.2)		
Time helping patients navigate (minutes)			0.0006[a] (1)	.98			0.51 (1)	.47
≤4	131/154 (85.1)	69/81 (85.2)			30/62 (49.4)	13/31 (40.6)		
≥5	23/154 (14.9)	12/81 (14.8)			32/62 (51.6)	19/31 (59.4)		
Ease of interpreter services (compared to in-person visits)			0.02 (1)	.90			1.16 (1)	.28
Somewhat or much more difficult	88/134 (65.7)	34/51 (66.7)			24/31 (77.4)	17/19 (89.5)		
Somewhat or much easier	46/134 (34.3)	17/51 (33.3)			7/31 (22.6)	2/19 (10.5)		
Audio and video quality			1.41 (1)	.24		N/A[b]	N/A	N/A
Not adequate	17/161 (10.6)	12/75 (16.0)			8/59 (13.6)	N/A	N/A	N/A
Adequate	144/161 89.4)	63/75 (84.0)			51/59 (86.4)	N/A	N/A	N/A
Device			4.24 (1)	.04[a]			3.31 (1)	.07
Only institution-provided devices	45/162 (27.8)	32/78 (41.0)			26/63 (41.3)	51/177 (28.8)		
At least one personal device used	117/162 (72.2)[a]	46/78 (59.0)			37/63 (58.7)	126/177 (71.2)		

[a]Significant at P<.05.

[b]N/A: not applicable.

Figure 2. Clinician-identified training needs for conduct of telemedicine, comparing adopters to low/nonadopters. Experience with 4 completed telephone visits in a half day session implies a high telephone adopter, and <4 visits implies a low telephone adopter. Experience with 1 video visit per half day on average implies a video visit adopter. *A higher proportion of video nonadopters stated a desire for training on the technical aspects of a telemedicine visit, compared to video adopters (49.6%, 57/115 vs 27.2%, 28/103; *P*<.001).
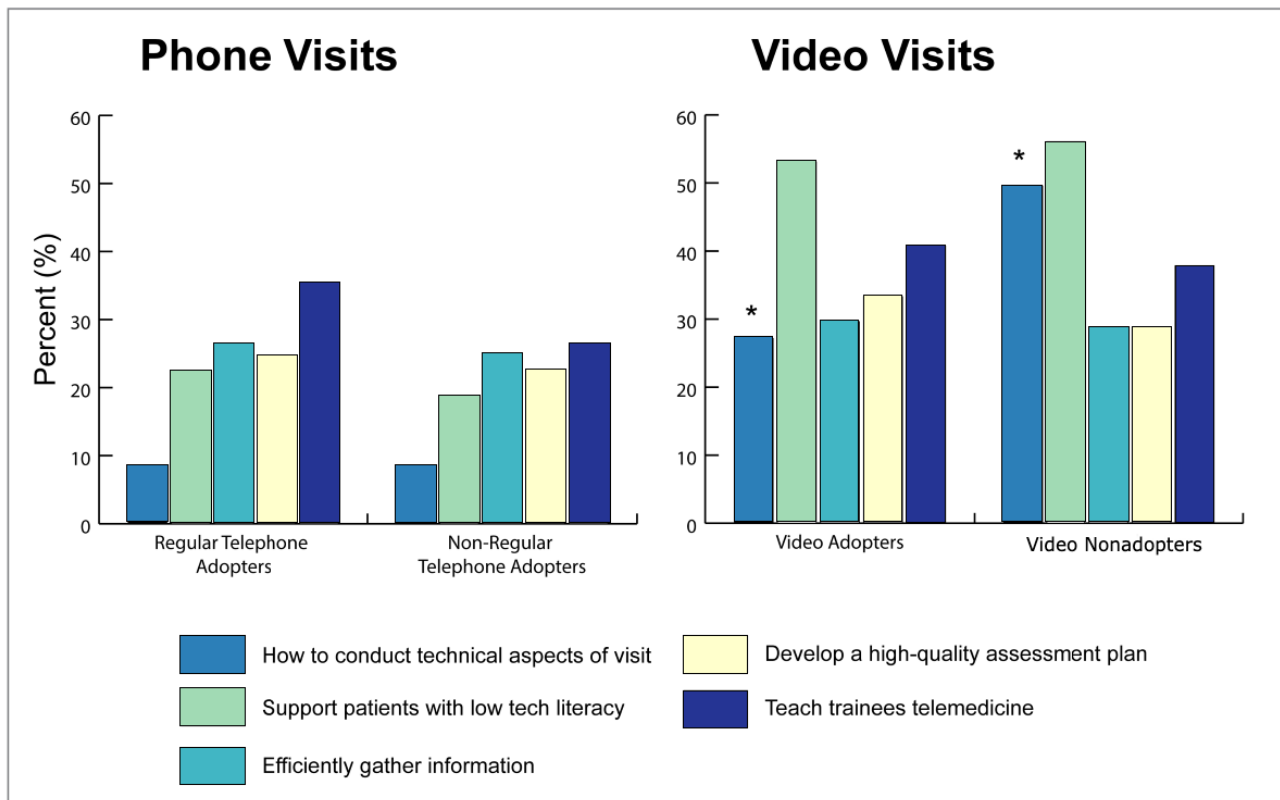
Discussion

Principal Findings

Our investigation is one of the few studies assessing system-level factors that might influence multispecialty clinicians' telephone and video visit adoption at a safety-net site in the early phase of the COVID-19 pandemic. Like many safety-net clinics, our telemedicine services were largely telephone-based [8].

We found that many clinicians utilize their personal devices to provide both telephone and video visits. A study of family planning clinicians found high personal smartphone use for telemedicine care during the COVID-19 pandemic, despite respondents' preference for work-issued devices [28]. This raises equity concerns, as not all clinicians will have the same access to high-quality video and audio via their smartphone, tablet device, or computer. In our network, departments variably distributed laptops with video capacity. Privacy and security are concerns if personal devices are used for encounters without appropriate encryption, 2-factor authentication, or HIPAA (Health Insurance Portability and Accountability Act)–approved software [29]. Our system supports a server-based thin client computing environment where most computers are run from a central server. Widespread use of hardware peripherals such as webcams for video visits on such devices increases risk for network instability. This will limit safety-net clinicians who practice on site to reliably use video visits without using or compromising the security of their personal devices [30,31].

We also observed higher personal device use among video nonadopters than video adopters, although this was not significant. This may be due to clinician reluctance to access clinical or HIPAA-relevant video through their personal device, which may be an additional barrier to video visit adoption that warrants further investigation.

Desire for more training in telemedicine was high for video visits, with almost half of video nonadopters requesting training in the conduct of the technical aspects of the visit. The extent and capacity to which clinicians have been trained in telemedicine modalities have been variable throughout the COVID-19 pandemic, and standardized trainings may influence adoption [32].

Although we could not statistically compare telephone adopters against video adopters, we see that barriers such as increased workload, greater time spent supporting the patient, difficulty with the interpreter, and poor audiovisual quality were more common for video adopters than high telephone adopters. In this network, clinicians must provide their own time and support necessary to allow patients to engage in video visits. Direct patient supports, through patient orientations or external services assisting patients in logging into video visits and contacting an interpreter, would drastically reduce these barriers.

Neither perceived workload of telemedicine (compared to in-person care), nor length of time spent supporting a patient to enter a telemedicine visit were associated with telemedicine adoption. Interpreter challenges were not associated with adoption, although this has been endorsed as a barrier by

clinicians [33]. Although safety-net patients may experience access barriers to bandwidth or adequate audio, audio quality was not associated with telemedicine adoption. Our sample size may have been too small to detect a relationship with these factors. Another interpretation is that clinicians using telemedicine during the COVID-19 pandemic were willing to surmount significant patient-level barriers to provide care. As telemedicine continues, it will be important to assess if later adopters will have different responses to these workload-related challenges.

Within this network, there was wide variation in infrastructural support for video visits by department. Some surgical specialties may have needed in-person evaluation rather than virtual visits. These factors likely explain specialty-specific differences in our findings.

Since our survey, telemedicine has become integrated into ambulatory care and is likely "here to stay," both owing to ongoing pandemic surges because of new variants, as well as patient and provider preference for the convenience and access enabled by virtual visits. For example, in this network, telemedicine volume appears to be approaching a plateau of 20%-30% of primary and specialty care, although video visit uptake remains low (internal data). Learning health systems serving vulnerable populations must assess the first years of implementation to optimize video visit delivery for the long term. Without system-level investments, disparities in video visit access may worsen disparities in the safety-net [34]. First, safety-net leaders should assess telemedicine utilization patterns by department to perceive variation in audio-only versus video visits. Reliance on personal device use should be surveyed; future research should query what devices are key for high-quality telemedicine interactions [35,36]. Third, further research should explore facilitators and barriers to video visits in safety-net settings, and identify what skills and trainings are most helpful to support telemedicine practice. Clinician ability to provide video visits will remain a priority in a shifting reimbursement policy landscape.

Limitations

We almost reached a 50% response rate for this voluntary survey among diverse clinicians. Although this was greater than our prespecified target, participants may have differed in their experience of telemedicine compared to nonparticipants. There was high variation in the completeness of survey responses; this was likely owing to the survey burden early during the pandemic. However, we observed no differences in results when comparing all participants to complete case analysis. Owing to survey missingness, we could not generate a multivariable model nor assess for variable interactions. As findings are based on bivariate analysis, they are exploratory and hypothesis-generating for system-level factors that may be related to safety-net telemedicine implementation. As telemedicine workflows were evolving during the survey period, some telemedicine practices may have changed during our assessment; this may have confounded some of our findings regarding specialty-specific differences. However, although a specific workflow was disseminated in June 2020, the same video visit software was available to clinicians over the entire survey period. While our definitions of "high telephone adoption" and "video adoption" were appropriate when we conducted the survey, telemedicine volume will continue to shift and evolve. Comparing the proportion of telemedicine visits to that of in-person visits may be an alternative definition for future studies; qualitative work should also assess how varying specialties choose to optimize their ambulatory schedules between virtual and in-person visits.

Conclusions

Clinical specialty type, personal device use, and desire for technical training were major factors associated with telephone and video visit adoption among safety-net clinicians. Department-level support, assessment of use of personal devices, and clinician training are priorities for safety-net systems.

Multimedia Appendix 1
Survey instrument.
[DOCX File , 56 KB - formative_v6i3e34088_app1.docx]

Multimedia Appendix 2
Respondent reported clinical specialities.

XSL•FO
RenderX

[DOCX File , 31 KB - formative_v6i3e34088_app2.docx]

References

1. New and expanded flexibilities for Rural Health Clinics (RHCs) and Federally Qualified Health Centers (FQHs) during the COVID-19 Public Health Emergency (PHE) Internet. Centers for Medicare and Medicaid Services. URL: https://www.cms.gov/regulations-and-guidanceguidancetransmittals2020-transmittals/se20016 [accessed 2020-10-15]

2. Lightbourne W, Newsom G. Medi-Cal Payment for Telehealth and Virtual/Telephonic Communications Relative to the 2019-Novel Coronavirus (COVID-19). Department of Health Care Services. URL: https://www.dhcs.ca.gov/Documents/COVID-19/Telehealth-Other-Virtual-Telephonic-Communications.pdf [accessed 2022-02-25]

3. Verma S. Early Impact Of CMS Expansion Of Medicare Telehealth During COVID-19 Internet. Health Affairs Blog. 2020 Jul 15. URL: https://www.healthaffairs.org/do/10.1377/hblog20200715.454789/full/ [accessed 2021-04-19]

4. Institute of Medicine (US) Committee on the Changing Market, Managed Care, and the Future Viability of Safety Net Providers. In: Lewin ME, Altman S, editors. Americas's Health Care Safety Net: Intact but Endangered. Washington, DC: National Academies Press; 2000.

5. Nguyen OK, Makam AN, Halm EA. National Use of Safety-Net Clinics for Primary Care among Adults with Non-Medicaid Insurance in the United States. PLoS One 2016;11(3):e0151610 [FREE Full text] [doi: 10.1371/journal.pone.0151610] [Medline: 27027617]

6. Kim J, Desai E, Cole M. How the rapid shift to telehealth leaves many community health centers behind during the COVID-19 pandemic Internet. Health Affairs. 2020 Jun 02. URL: https://www.healthaffairs.org/do/10.1377/hblog20200529.449762/full/ [accessed 2020-10-16]

7. Lau J, Knudsen J, Jackson H, Wallach AB, Bouton M, Natsui S, et al. Staying Connected In The COVID-19 Pandemic: Telehealth At The Largest Safety-Net System In The United States. Health Aff (Millwood) 2020 Aug;39(8):1437-1442. [doi: 10.1377/hlthaff.2020.00903] [Medline: 32525705]

8. Uscher-Pines L, Sousa J, Jones M, Whaley C, Perrone C, McCullough C, et al. Telehealth Use Among Safety-Net Organizations in California During the COVID-19 Pandemic. JAMA 2021 Mar 16;325(11):1106-1107 [FREE Full text] [doi: 10.1001/jama.2021.0282] [Medline: 33528494]

9. Serper M, Nunes F, Ahmad N, Roberts D, Metz DC, Mehta SJ. Positive Early Patient and Clinician Experience with Telemedicine in an Academic Gastroenterology Practice During the COVID-19 Pandemic. Gastroenterology 2020 Oct;159(4):1589-1591.e4 [FREE Full text] [doi: 10.1053/j.gastro.2020.06.034] [Medline: 32565015]

10. Audio-Only Delivery. Center for Connected Health Policy. URL: https://www.cchpca.org/topic/audio-only-delivery-covid-19/ [accessed 2022-01-24]

11. Post-COVID-19 Public Health Emergency Telehealth Policy Recommendations: Public Document. State of California—Health and Human Services Agency. Department of Health Care Services. 2021 Feb 02. URL: https://www.dhcs.ca.gov/services/medi-cal/Documents/DHCS-Telehealth-Policy-Proposal-2-1-21.pdf [accessed 2021-05-07]

12. Proposed Policy, Payment, and Quality Provisions Changes to the Medicare Physician Fee Schedule for Calendar Year 2021. Centers for Medicare & Medicaid Services. 2020 Aug 03. URL: https://www.cms.gov/newsroom/fact-sheets/proposed-policy-payment-and-quality-provisions-changes-medicare-physician-fee-schedule-calendar-year-4 [accessed 2021-04-27]

13. Eberly LA, Khatana SAM, Nathan AS, Snider C, Julien HM, Deleener ME, et al. Telemedicine Outpatient Cardiovascular Care During the COVID-19 Pandemic: Bridging or Opening the Digital Divide? Circulation 2020 Aug 04;142(5):510-512 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.120.048185] [Medline: 32510987]

14. Reed ME, Huang J, Graetz I, Lee C, Muelly E, Kennedy C, et al. Patient Characteristics Associated With Choosing a Telemedicine Visit vs Office Visit With the Same Primary Care Clinicians. JAMA Netw Open 2020 Jun 01;3(6):e205873 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.5873] [Medline: 32585018]

15. Schifeling CH, Shanbhag P, Johnson A, Atwater RC, Koljack C, Parnes BL, et al. Disparities in Video and Telephone Visits Among Older Adults During the COVID-19 Pandemic: Cross-Sectional Analysis. JMIR Aging 2020 Nov 10;3(2):e23176 [FREE Full text] [doi: 10.2196/23176] [Medline: 33048821]

16. Ye S, Kronish I, Fleck E, Fleischut P, Homma S, Masini D, et al. Telemedicine Expansion During the COVID-19 Pandemic and the Potential for Technology-Driven Disparities. J Gen Intern Med 2021 Jan;36(1):256-258 [FREE Full text] [doi: 10.1007/s11606-020-06322-y] [Medline: 33105000]

17. Nouri S, Khoong E, Lyles C, Karliner L. Addressing equity in telemedicine for chronic disease management during the covid-19 pandemic. NEJM Catalyst Innovations in Care Delivery 2020 May 04 [FREE Full text] [doi: 10.1056/CAT.20.0123]

18. Rodriguez JA, Betancourt JR, Sequist TD, Ganguli I. Differences in the use of telephone and video telemedicine visits during the COVID-19 pandemic. Am J Manag Care 2021 Jan;27(1):21-26 [FREE Full text] [doi: 10.37765/ajmc.2021.88573] [Medline: 33471458]

19. Haque SN, DeStefano S, Banger A, Rutledge R, Romaire M. Factors Influencing Telehealth Implementation and Use in Frontier Critical Access Hospitals: Qualitative Study. JMIR Form Res 2021 May 05;5(5):e24118 [FREE Full text] [doi: 10.2196/24118] [Medline: 33949958]

20. Khoong EC, Butler BA, Mesina O, Su G, DeFries TB, Nijagal M, et al. Patient interest in and barriers to telemedicine video visits in a multilingual urban safety-net system. J Am Med Inform Assoc 2021 Feb 15;28(2):349-353 [FREE Full text] [doi: 10.1093/jamia/ocaa234] [Medline: 33164063]

21. Sadasivaiah S, Shaffer E, Enanoria W, Su G, Goldman S, Scarafia J, et al. Informatics response to address the COVID-19 pandemic in a safety net healthcare system. JAMIA Open 2021 Jul;4(3):ooaa057 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa057] [Medline: 34504999]

22. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]

23. Langbecker D, Caffery LJ, Gillespie N, Smith AC. Using survey methods in telehealth research: A practical guide. J Telemed Telecare 2017 Oct;23(9):770-779. [doi: 10.1177/1357633X17721814] [Medline: 28728502]

24. Parmanto B, Lewis AN, Graham KM, Bertolet MH. Development of the Telehealth Usability Questionnaire (TUQ). Int J Telerehabil 2016;8(1):3-10 [FREE Full text] [doi: 10.5195/ijt.2016.6196] [Medline: 27563386]

25. Venkatesh V, Davis FD. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Manage Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]

26. Lyles C, Fields J, Lisker S, Sharma A, Aulakh V, Sarkar U. Launching a Toolkit for Safety-Net Clinics Implementing Telemedicine During the COVID-19 Pandemic Internet. The Common Health Fund. URL: https://www. commonwealthfund.org/blog/2020/launching-toolkit-safety-net-clinics-implementing-telemedicine-during-covid-19-pandemic [accessed 2021-04-19]

27. Martinez M, San FD. Principles for Collecting, Coding,and ReportingIdentity Data Sexual Orientation Guidelines Internet. San Francisco Department of Public Health. 2014. URL: https://www.sfdph.org/dph/files/PoliciesProcedures/ COM9_SexualOrientationGuidelines.pdf [accessed 2021-04-19]

28. Stifani BM, Avila K, Levi EE. Telemedicine for contraceptive counseling: An exploratory survey of US family planning providers following rapid adoption of services during the COVID-19 pandemic. Contraception 2021 Mar;103(3):157-162. [doi: 10.1016/j.contraception.2020.11.006] [Medline: 33212033]

29. Smith KA, Zhou L, Watzlaf VJM. User Authentication in Smartphones for Telehealth. Int J Telerehabil 2017;9(2):3-12 [FREE Full text] [doi: 10.5195/ijt.2017.6226] [Medline: 29238444]

30. Al Ayubi SU, Pelletier A, Sunthara G, Gujral N, Mittal V, Bourgeois FC. A Mobile App Development Guideline for Hospital Settings: Maximizing the Use of and Minimizing the Security Risks of "Bring Your Own Devices" Policies. JMIR Mhealth Uhealth 2016 May 11;4(2):e50 [FREE Full text] [doi: 10.2196/mhealth.4424] [Medline: 27169345]

31. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Privacy and security in mobile health apps: a review and recommendations. J Med Syst 2015 Jan;39(1):181. [doi: 10.1007/s10916-014-0181-3] [Medline: 25486895]

32. Meyer B, Friedman L, Payne K, Moore L, Cressler J, Holberg S, et al. Medical Undistancing Through Telemedicine: A Model Enabling Rapid Telemedicine Deployment in an Academic Health Center During the COVID-19 Pandemic. Telemed J E Health 2021 Jun;27(6):625-634. [doi: 10.1089/tmj.2020.0327] [Medline: 33030985]

33. Sharma AE, Khoong EC, Nijagal MA, Lyles CR, Su G, DeFries T, et al. Clinician Experience with Telemedicine at a Safety-net Hospital Network during COVID-19: A Cross-sectional Survey. J Health Care Poor Underserved 2021;32(2S):220-240. [doi: 10.1353/hpu.2021.0060]

34. Uscher-Pines L, Schulson L. Rethinking The Impact Of Audio-Only Visits On Health Equity. Health Affairs Blog. 2021 Dec 17. URL: https://www.healthaffairs.org/do/10.1377/forefront.20211215.549778/full/ [accessed 2022-01-24]

35. Kadimo K, Kebaetse MB, Ketshogileng D, Seru LE, Sebina KB, Kovarik C, et al. Bring-your-own-device in medical schools and healthcare facilities: A review of the literature. Int J Med Inform 2018 Nov;119:94-102. [doi: 10.1016/j.ijmedinf.2018.09.013] [Medline: 30342692]

36. Wani TA, Mendoza A, Gray K. Hospital Bring-Your-Own-Device Security Challenges and Solutions: Systematic Review of Gray Literature. JMIR Mhealth Uhealth 2020 Jun 18;8(6):e18175 [FREE Full text] [doi: 10.2196/18175] [Medline: 32554388]

Abbreviations

EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act

XSL•FO
RenderX

Original Paper

COVID-19 Mental Health Stressors of Health Care Providers in the Pandemic Acceptance and Commitment to Empowerment Response (PACER) Intervention: Qualitative Study

Christa Sato[1], BA, BSW, MSW; Anita Adumattah[2], MN, PHCNP; Maria Krisel Abulencia[2], MN; Peter Dennis Garcellano[2], MN; Alan Tai-Wai Li[3,4], MD; Kenneth Fung[5], MD; Maurice Kwong-Lai Poon[6], MSW, PhD; Mandana Vahabi[2], PhD; Josephine Pui-Hing Wong[2], PhD

[1]Factor-Inwentash Faculty of Social Work, University of Toronto, Toronto, ON, Canada
[2]Daphne Cockwell School of Nursing, Ryerson University, Toronto, ON, Canada
[3]Regent Park Community Health Centre, Toronto, ON, Canada
[4]Committee for Accessible AIDS Treatment, Toronto, ON, Canada
[5]Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[6]School of Social Work, York University, Toronto, ON, Canada

Corresponding Author:
Josephine Pui-Hing Wong, PhD
Daphne Cockwell School of Nursing
Ryerson University
350 Victoria Street
Toronto, ON, M5B2K3
Canada
Phone: 1 416 979 5000 ext 556303
Email: jph.wong@ryerson.ca

Abstract

Background: Since the pandemic, more Canadians have reported poorer mental health. A vital group experiencing a high level of stressors consists of health care providers (HCPs) caring for COVID-19 patients, carrying out public health responses, or working with vulnerable populations. The mental health of HCPs is negatively affected by the pandemic, not only at work but also at home and in the community. Intersecting stressors at multiple levels contribute to HCPs' experiences of fatigue, insomnia, anxiety, depression, and posttraumatic stress symptoms.

Objective: The aim of this qualitative study was to explore the pandemic stressors experienced by HCPs at work, at home, and in the community before participating in the Pandemic Acceptance and Commitment to Empowerment Response (PACER) online intervention.

Methods: Informed by a social ecological approach, we used a qualitative reflective approach to engage 74 HCPs in diverse roles. Data were collected during the first 2 waves of the COVID-19 pandemic (June 2020 to February 2021) in Canada.

Results: Informed by a social ecological framework, 5 overarching themes were identified in our thematic analysis: (1) personal level stressors that highlight HCPs' identities and responsibilities beyond the workplace; (2) interpersonal level stressors from disrupted social relationships; (3) organizational stressors that contributed to unsettled workplaces and moral distress; (4) community and societal stressors attributed to vicarious trauma and emotional labor; and (5) the multilevel and cumulative impacts of COVID-19 stressors on HCPs' health.

Conclusions: COVID-19 is not merely a communicable disease but also a social and political phenomenon that intensifies the effects of social inequities. Current understanding of pandemic stressors affecting HCPs is largely partial in nature. Although workplace stressors of HCPs are real and intense, they need to be explored and understood in the context of stressors that exist in other domains of HCPs' lives such as family and community to ensure these experiences are not being silenced by the "hero" discourses or overshadowed by professional demands.

XSL•FO
RenderX

Introduction

In Canada, the COVID-19 pandemic has claimed over 31,000 lives [1]. It has profoundly affected the lives of Canadians, particularly their mental health and well-being. Although everyone is affected by the pandemic, the degree of COVID-19 impact varies across groups and communities. Since the pandemic, more Canadians have reported poorer mental health, particularly women, individuals with chronic illnesses or disabilities, racialized persons, immigrants, and Indigenous peoples [2,3]. A vital group experiencing a high level of stressors consists of health care providers (HCPs) directly involved in caring for COVID-19 patients, carrying out public health responses, or working with vulnerable populations. Emerging evidence in Canada and elsewhere indicates that the mental health of HCPs is negatively affected by the pandemic. HCPs have reported sleep disturbances, fatigue, and episodes of insomnia [4,5] as well as increased prevalence of anxiety [6,7], depressive symptoms [8,9], suicide ideation [8], and posttraumatic stress symptoms triggered by repeated intense stressful experiences and reduced social life [5,10].

During a pandemic, the popular image of HCPs is often limited to individuals in personal protective equipment (PPE) and saving lives. However, behind the masks and full PPE gear, HCPs are people with many social identities, roles, relationships, responsibilities, emotions, and needs. There is relatively limited qualitative research on the experiences of stress and challenges that negatively affect the mental health of Canadian HCPs, particularly those that extend beyond their professional identities and roles. A study with 20 Canadian nurses identified uncertainty, inadequate access to PPE, repeated witnessing of patient deterioration and death, and experience with social isolation as key sources of stress [11]. Two other studies found that HCPs experience increased stressors related to rapidly changing policy, unclear or lack of leadership communication, anxiety about safety for self and family, and tension between meeting the needs of their patients and pandemic protocols [9,12]. Studies in other countries also revealed similar multilevel stressors that affect HCPs' mental health. Additional stressors include caring for and losing one's own infected colleagues to COVID-19, inconsistent guidelines on COVID-19 testing and treatment, ineffective staff redeployment, and staff shortages [13-15]. All these intersecting challenges also contribute to moral distress, which can be defined as the emotional, psychological, and physical suffering that HCPs experience when they carry out their responsibilities in ways that are inconsistent or contradictory to the ethical values, principles, or commitments that they hold deeply [16].

Current understandings of pandemic stressors affecting HCPs are largely partial in nature. Since the beginning of the COVID-19 pandemic, HCPs (especially nurses) have been hailed as "unsung heroes" by politicians, the mass media, and the general public. In their poststructural analysis of over 70 documents in mass media and social media on "nurses as heroes," Mohammed and colleagues [17] argued that the hero discourse positions nursing as "model citizens" and "a necessary sacrifice," whereby heroism is the reward. When taken together, the hero discourse imposes a new form of subjectivity and normalizes nurses' exposure to workplace risks. We further argue that the hero discourse has created a spotlight that solely focuses on the professional identity and roles of HCPs, negating the multiple identities that constitute their "holistic self" and their everyday lived experiences that are inseparable and indistinguishable in their intertwined home-work-community life. Thus, while workplace stressors of HCPs are real and intense, they need to be explored and understood in the context of stressors that exist in other domains of HCPs' lives. In the context of a pandemic, stressors can be understood as an ever-present condition that threatens the health or well-being of individuals [18].

To date, an overwhelming majority of empirical studies on pandemic-related stressors among HCPs are quantitative in nature and report specific measurable aspects of stressors that are often decontextualized. Qualitative research on this topic is needed to provide insights on the multilevel contexts of how HCPs experience and are impacted by COVID-19 stressors. In this paper, we report on the qualitative data collected from HCPs who self-identified as experiencing a high level of stress and wished to enroll in an online stress-reduction intervention designed for HCPs and affected communities. The results presented contribute to addressing the gaps in our current understanding of pandemic stressors experienced by HCPs in the Canadian context.

Methods

Study Setting: The PACER Intervention

The Pandemic Acceptance and Commitment to Empowerment Response (PACER) intervention [19] is an evidence-informed 6-week online training based on the Acceptance and Commitment to Empowerment (ACE) model. ACE is an innovative integration of the mindfulness-based Acceptance and Commitment Therapy (ACT) [20] and the social justice-based Group Empowerment Psychoeducation (GEP) [21]. It has been applied effectively in previous in-person interventions to reduce stigma and promote resilience. Recognizing the growing mental health needs of frontline HCPs and constrained by pandemic lockdowns, we developed the PACER intervention, which consists of 6 weekly online self-guided learning modules, each lasting 1 hour, accompanied by weekly 90-minute facilitated online group video-conferencing sessions. Prior to engaging in the first online learning module and first group sessions, participants were invited to complete a sociodemographic survey and questionnaires that consisted of both quantitative and qualitative questions (see the Procedure section).

Ethical Review

This study was reviewed and approved by the Research Ethics Boards at Ryerson University (2020-123), University Health Network (20-5210), and York University (2020-096).

Participants

We used convenience and snowball sampling, supported by targeted outreach and recruitment via circulations of e-flyers through our collaborators, nursing professional associations, hospital bulletins, community networks, and social media. The participation criteria were self-identifying as 18 years of age or older, being a service provider in the health care sector, experiencing high levels of stress related to the COVID-19 response, and interested in joining the PACER training. HCP participants in this study included registered nurses (RNs), registered practical nurses (RPNs), personal support workers, physicians, medical residents, social workers, counsellors, therapists, and other clinical or community-based personnel.

A total of 74 HCPs enrolled and took part in PACER between June 2020 and February 2021: 63 (85%) self-identified as female, 10 (14%) self-identified as male, and 1 (1%) self-identified as transgender. Most participants were under 40 years of age: 20-29 years: 10/74, 14%; 30-39 years: 38/74, 51%; 40-49 years: 14/74, 19%; 50-59 years: 9/74, 12%; and 60-69 years: 3/74, 4%. When asked about the ethnocultural communities they identified with, about one-half of the participants (38/74, 51%) indicated that they identified with the Canadian community, while the other one-half (36/74, 49%) identified with multiple communities including African, American, Chinese, Filipino, Haitian, Irish, Hong Kong, Taiwan, Jewish, and Latinx. Participants came from a wide range of health and social care professional backgrounds and roles such as primary care, psychiatry, nursing, social work, public health, community health practitioners, therapists/counsellors, administrative staff, environmental service staff, and volunteers.

Procedure

Before data collection, online informed consent was obtained from all participants. Once enrolled, participants were invited to complete a pre-intervention questionnaire that consisted of sociodemographic questions and measures on general mental health distress, psychological flexibility, and resilience. These measures were repeated immediately post- and 3 months postintervention. In addition, prior to Session One of the intervention, we used qualitative open-ended questions to explore participants' self-identities and experiences with pandemic stressors. Specifically, we invited participants to describe "who you are" and "the impact of COVID-19 on your work and your life—what has been challenging for you."

Analysis

The focus of this paper was on the pandemic-related stressors reported by participants in the qualitative questions before they engaged in the 6-week online intervention and online group sessions. Responses from the online questionnaires were downloaded, and pseudonyms are used in this paper to protect participants' identities. Our thematic analysis was informed by the social ecological framework [22], which allowed us to examine the sources of stressors experienced by HCPs at multiple interconnected levels [23]: personal (identity, social position), interpersonal (family, relationships, social networks), organizational (operational characteristics, formal and informal rules), community contexts (relationships among organizations and institutions), and societal (local, provincial, and national guidelines and policies) [24]. This approach enabled us to make sense of how different synergistic factors and contexts intersect to shape the experiences with pandemic stressors among HCPs.

The first 4 authors, under the guidance of the last author, conducted thematic analysis [25] using both inductive and deductive approaches. First, we engaged in repeated readings of the participant responses to gain a broad understanding of the data. Second, we identified and developed thematic codes inductively based on the ideas and perspectives articulated by the participants (ie, data-based coding). Third, guided by our research purpose and the social ecological framework, we applied deductive analysis that enabled us to connect the data to the sociocultural contexts and structural conditions that shaped the experiences of the participants [26,27]. Fourth, the team engaged in multiple rounds of interpretive discussions to arrive at the agreed themes presented in this paper. In the last round of thematic interpretation, all authors took part to discuss the nuances and contexts of the narratives under each theme to arrive at a holistic interpretation.

Results

Informed by a social ecological framework, 5 overarching themes were identified in our thematic analysis: (1) personal-level stressors that highlight HCPs' identities and responsibilities beyond the workplace, (2) interpersonal-level stressors from disrupted social relationships, (3) organizational stressors that contributed to unsettled workplaces and moral distress, (4) community and societal stressors attributed to vicarious trauma and emotional labor, and (5) the multilevel and cumulative impacts of COVID-19 stressors on HCPs' health.

Personal Stressors: Identities and Responsibilities Beyond the Workplace

Stressors at the personal level emphasize the identities and responsibilities of HCPs above and beyond the workplace that participants encountered in their day-to-day lives. Many participants expressed fear and worries about bringing the SARS-CoV-2 virus home to their elderly family members or loved ones with existing health problems.

> *My husband had been undergoing intensive chemotherapy and radiation treatment at [a cancer care hospital], and was halfway through a year of treatment... I was very scared that I might contract the virus at work and bring it back to our house... and impact his access to his next treatment.* [Sharon, 30-39 years, female, children's rehab coordinator, Canadian]

It is also important to note that HCPs are not only service providers; some relied on accessing services to meet their own health needs or the needs of their family members. Hence, service disruptions are not only felt professionally but also personally at home or in their communities.

As the parent of a child with dual diagnosis I already feel isolated from society. My son is not capable of wearing a mask and as a result we have not been to any activities together since the first lockdown... My son does not understand the restrictions and has been suffering from lack of social interaction and engagement in his community. [Kathy, 30-39 years, female, behavior therapist, Canadian]

To meet the demands of their multiple roles and responsibilities, some participants had to make use of available resources to mitigate some of these struggles. Sun, an RN, had to rely on her elderly in-laws to care for her toddler while she was at work. Meanwhile, she also expressed her concerns:

[We] fear the potential transmission of COVID-19 to my in-laws as they are of higher risk from pre-existing medical conditions and advanced age." [Sun, female, 30-39 years, RN and case manager, Chinese-Canadian, Hong Kong]

Sun's fear was echoed by other HCPs living in multigenerational households comprised of vulnerable elderly adults and young children.

Interpersonal Stressors: Disrupted Social Relationships

Similar to all Canadians, many participants identified disruptions to routines and structures in their personal life as stressors. Social distancing and quarantine measures had heightened participants' sense of isolation and loneliness where social relationships are vital resources in fostering HCPs' resiliency. For participants with many demands and responsibilities, travelling is not merely a leisure activity but a coping strategy to reduce stress. Andrew, who had to manage demands from school, work, and family, indicated that his once-a-year vacation was "a breather" from his stressful life. He further explained:

The vacation time is my only chance to see my family, who lives in the Philippines. As the pandemic unfolds, all my plans are put on hold." [Andrew, male, 30-39 years, RPN in retirement facility, Filipino-Chinese]

For participants with transnational family and social ties, pandemic travel restrictions resulted in prolonged separation and disconnection from their families. Those with elderly or sick family members felt additional challenges.

The biggest issue was, my mother is in advanced Alzheimer's back in India and her health had started to decline. Prior to [the] COVID-19 outbreak, I had booked a flight for India, which of course got cancelled. [Priya, female, 40-49 years, RPN in geriatric care facility and BScN student, South Asian]

Furthermore, while the pandemic had forced many participants to put their plans on hold, it had not stopped other critical events from taking place around the world. Participants with families and loved ones in Iran, India, and the United States expressed a sense of powerlessness and worries due to the pandemic situation worldwide.

I have felt very helpless while watching the effects of COVID tear through the U.S., on top of all the other crises that the country is currently facing. I spend a lot of time worrying about my parents and I have had difficulty focusing on my schoolwork. [Adam, male, 20-29 years, graduate student in clinical psychology, American]

Participants' sharing illustrated that, in addition to their demands as frontline HCPs, they also shouldered many demands similar to the general population, even though these stressors were overshadowed by discourses of their professional demands.

Organizational Stressors: Unsettled Workplace and Moral Distress

Not surprisingly, HCP participants encountered many workplace stressors, ranging from increased health risks related to direct exposure to COVID-19, changes in work routines and workloads, and the need to wear PPE all day. At the beginning of the pandemic, uncertainty about COVID-19 created a lot of fear among HCPs, particularly those with less formal biomedical or health care training. Matthew, an environment service staff member in a hospital, expressed tremendous fear toward COVID-19.

When I was assigned to clean rooms where there were patients with COVID-19... I was too scared to communicate with the patients. I limit our conversation by just answering yes or no. I feel very guilty because of my fear... I am working again tomorrow, I hope I will not encounter a patient with COVID." [Matthew, male, 30-39 years, environmental service staff, Filipino]

For some HCPs, their fear was not limited to personal safety but extended to the uncertainty and demands of keeping their staff and patients or vulnerable service users safe.

At times I feel overwhelmed that my support is not sufficient in supporting a large team. Stressed by the ongoing changes, angry at the impact it's having on the homeless community and the lack of acknowledgement frontline staff in shelters and homelessness services receive. [Tracy, female, 40-49 years, community practitioner, Canadian]

The rapid changes in the workplace routines and structures also led to considerable distress that altered or strained participants' interpersonal relationships with their patients, colleagues, and other staff. For Kathy, a behavior therapist working with children diagnosed with autism spectrum disorder, it was difficult to carry out her work with PPE:

I miss being able to high five my clients and look into their eyes unobstructed by PPE in order to read the small changes in their expressions. [Kathy, female, 30-39 years, behavior therapist, Canadian]

Other HCPs were concerned about access to safe PPE due to previous experience of receiving defective PPE.

Many participants identified deployment as a major source of stress. One RN experienced anger and resentment "when colleagues were redeployed to other areas of the hospital," but her team was "expected to maintain the same caseload with less human resources" [Sun, female, 30-39 years, RN, Chinese-Canadian, Hong Kong].

Marianna, who was deployed to work in nursing homes, recalled the stress and exhaustion:

> It was a bit stressful to learn 3 different positions within 3 months... I often worked overtime and weekends, so it left long stretches of very little time for myself and my family. [Marianna, female, 40-49 years, RN, Portuguese Canadian]

In addition to deployment and increased workload, some HCPs experienced moral distress when their efforts to provide safe and quality care to their vulnerable patients were compromised by the rapidly changing policies at work, as Giselle explained:

> Most of our patients have a history with mental health challenges and trauma. While my workplace is based on a harm reduction philosophy, it felt like we were also re-traumatizing our patients due to the new COVID-related policies. [Giselle, female, 20-29 years, RN, Filipino]

Similarly, Ariana expressed her moral struggles and dilemma in attending to the special needs of families in her clinic.

> Families... having their kids with behavior and mood challenges at home with them, much reduced access to in-person and out-of-home respite and other services... I am the point person for families... Ethical dilemma of having to medicate children to keep everyone sane and safe when we know it's the circumstances brought about by COVID-19 restrictions that is making things worse. [Ariana, female, 50-59 years, clinical nurse, Canadian, Welsh]

Other participants expressed concerns about workplace opacity and suggested that their employers had used pandemic emergency protocols to implement "different priorities" without consulting staff and patients. When participants' concerns about safety and accountability toward patient care were dismissed, they were left feeling undervalued and disempowered.

> When feedback was provided to management in regards to concerns about safe practice and lack of proper transfer of accountability between teams... I was met with ridicule and publicly humiliated, which caused me to feel insignificant, with feelings of powerlessness. [Sun, female, 30-39 years, RN and case manager, Chinese-Canadian, Hong Kong]

As the HCP participants encountered stressors of an unsettled workplace, many also struggled with the moral distress of weighing the quality of care they could realistically provide to patients against the safety of self and others. Other stressors included guilt for minimizing interactions with clients for fear of contracting the virus, the impact of services on staff and clients working with vulnerable populations, or feeling dismissed when concerns about workplace policies and practices are brought to the attention of administrators.

Community and Societal Stressors: Vicarious Trauma and Emotional Labor

Another stressor faced by HCPs during the pandemic is the intensified emotional labor placed on them at work. There seemed to be a collective resonance of vicarious trauma expressed by participants who witnessed the suffering of their patients or service users.

> Witnessing the impact of Covid-19 on seniors and feeling helpless has a negative impact on my own mental health. One lonely senior passed away and one senior's cat was missing, that especially hit hard on mental health. [Margaret, female, 50-59 years, community health promoter, Chinese-Canadian, Mainland China]

Amy, who witnessed and experienced the collective suffering, expressed her feelings of sadness:

> Clients are dying alone or with the touch of a gloved hand; loved ones are not able to say their goodbyes; and questions are not always answered... makes me feel drained emotionally and physically... and it has made me feel guilty. [Amy, female, 30-39 years, RN and NP student, African-Canadian]

As caregivers, HCPs often carried the emotional "burdens" of suppressing their own feelings while supporting and reassuring their loved ones, colleagues, patients, and service users who were in distress. Some of them also experience vicarious trauma as they witnessed the inequities and social oppressions experienced by individuals, families, and communities they worked with.

> It has been devastating to see the inequity and injustice around, who has been most impacted by the pandemic, and this is evident in the families I support. Community resources are stretched and limited so there are less options for support. This has resulted in higher demands on our service. All of this combined has left me feeling more depleted and less hopeful. [Julia, 40-49 years, social worker of children living with disabilities, Canadian]

Peggy, a Chinese-Canadian, described the pain in witnessing the most marginalized in the Asian Canadian community be further marginalized during the pandemic, due to "loss of income, mental health deterioration from social isolation, school closures, and experiences of anti-Asian racism" [female, 30-39 years, youth mental health clinician].

The demands of pandemic challenges were also felt by participants in management positions. Some of them felt inadequate when they were faced with concerns about the psychological and physical safety of both their staff and patients. This emotional labor could be difficult to bear when their efforts were met with abusive behaviors from the service users, as Louisa indicated:

> Over the course of the past several months, I have been subject to ongoing verbal abuse by patients and families, who disagree with the COVID-19 related precautions that have been implemented. [Louisa, female, 30-39 years, hospital clinical manager, Canadian]

Other HCPs reported experiencing overwhelming feelings of anger, frustration, worry, sadness, confusion, loneliness, helplessness, and burnout. Sun, who worked as a case manager, recalled:

...there was a period between April and June where I felt overwhelmed and started experiencing bitterness and resentment toward being a health care provider. [Sun, female, 30-39 years, RN, Chinese-Canadian, Hong Kong]

Despite the overwhelming emotional demands and sense of uncertainty, some participants were also inspired by a sense of collectivism. Many commented on the strength they received from sharing a strong sense of team spirit that brought value and meaning to their work.

The collaboration and teamwork! The all-hands-on-deck approach to my organization's response to the pandemic. It was not as smooth, but the teamwork was visible in ensuring that staff are aware of policy changes or updates. [Julian, male, 30-39 years, RN, Filipino-Canadian]

Other HCPs derived their sense of strength and optimism by focusing on "being part of a team working toward a greater goal, that we are all in this together, and a shared understanding of what it is to be a front-line worker during a pandemic" [Kamal, female, 30-39 years, intensive care unit nurse and critical care research coordinator, Canadian, Indian]. Hence, despite all the challenges encountered by HCPs, there were factors and contexts that promoted a sense of togetherness and connection with others, which can help foster resilience.

The Multilevel and Cumulative Impact of COVID-19 Stressors on Health

As highlighted in the previous themes, participants experienced multiple ongoing and compounding pandemic stressors in the domains of work, home, and society that undoubtedly impacted their physical, psychological, and emotional health and well-being. These impacts can also be considered stressors in and of themselves because dealing with the impact of these challenges simultaneously contributes to secondary stressors.

COVID-19 and the quarantine process has led to a collective trauma, even those whose physical health is functioning, but they are impacted emotionally... I have had moments of wanting to end my life back in April. [Jess, gender-fluid, 30-39 years, health discipline student, Canadian, Hong Kong]

Others described a range of physical problems from feeling drained, tired, burned out, concerns regarding sleep, poor concentration, and feeling tension and pain in their bodies. In some cases, individuals had pre-existing health conditions that worsened during the pandemic.

I am currently undergoing physiotherapy because I suffered a hip injury two years ago. Closure of the facility where I receive physiotherapy services [has] been a significant factor why I experience pain while I work. I only rely on my daily stretches and pain medications to cover-up the anticipated pain that I will experience at work. [Matthew, male, 30-39 years, environmental service staff, Filipino]

For some HCPs, deployment and workplace chaos had taken a toll on their mental health, yet they felt guilty and blamed themselves.

When I was redeployed to an entirely new department my anxiety intensified, to the point that I needed to seek accommodation to work from home. During this time I had negative thoughts about myself, including thoughts that I was weak, that I was letting my colleagues down, or that I should be able to handle this better. [Julia, 40-49 years, female, social worker in field of disabilities, Canadian]

Other HCPs prioritized the needs of others, sometimes to their own detriment because they neglected their own emotional and psychological needs.

I felt like I lived ten years. I felt like I was in a war for a long time. I stayed up till 3:30 in the morning for almost 2 months. I slept about 2 or 3 hours then went to work. I had to do 14 shifts straight without a day off due to staff shortage... I feel powerless to control my own life. [Lynn, 50-59 years, RN in long-term care (LTC) home, Chinese Canadian]

For Grace, the mother of an adult son with developmental disabilities, the pandemic made her feel anxious about the future of her son.

COVID has also highlighted painful thoughts about what would happen to my son if anything should happen to me... COVID has made me think more about my own mortality and how much work is still left to be done to prepare my son [who is an adult with developmental disabilities]. [Grace, female, 60-69 years, parent advisor in developmental disability sector, Canadian]

Participant responses in this section illustrate the multilevel and cumulative impact of pandemic stressors experienced by HCPs that were often intricately connected to pre-existing sociocultural contexts and structural conditions of their lives, as in caring for children with disabilities or managing new and existing psychological distress.

Discussion

Principal Findings

In this paper, we applied a social ecological approach to make visible the complex synergistic stressors experienced by HCPs at home, at work, and in the community. The main findings of the study highlighted 5 key pandemic-related stressors experienced by HCPs that manifested at the personal, interpersonal, organizational, community, and societal levels. Taken together, the findings contribute to a more holistic understanding of the stressors and impacts experienced by HCPs during the first 2 waves of the COVID-19 pandemic in Canada.

Many of these stressors were shaped by pre-existing structural conditions. Although many of the workplace-related stressors for HCPs (eg, fear of getting infected, anxiety related to uncertainty, rapid changes in protocols, deployment, staff shortage) identified in this study are consistent with what has been reported in the emerging literature and news reports [13-15], this study offers insights drawn from participants' qualitative reflections to contextualize and illustrate how workplace stressors interact with other stressors in the everyday

lives of HCPs to negatively affect their physical, emotional, and mental health. We also illustrate how the experiences of individual HCPs are nested in organizational practices, which are shaped by structural influences and power relations in society.

At the personal and interpersonal levels, most of the participants were also caregivers before and after work, as daughters and caregivers of elderly family members, mothers of young children, or caregivers of loved ones living with special needs, disabilities, or serious illnesses. Yet, their personal life was intricately connected to their professional life and shaped by structural conditions. With or without the pandemic, unpaid and unrecognized caregiving in both the public and private realm is well recognized to be "invisible labor" that is also gendered and racialized [28,29]. In 2020, women made up over 80% of all workers in health care and social services in Canada [30]. Racialized immigrants, particularly adults from the Caribbean, Africa, and Southeast Asia, are overrepresented in nursing and health care support occupations, and over 58% of these adult immigrants with international training were overqualified for their employment [30]. Reflective of the statistics on the Canadian health care workforce, 85% of our 74 HCP participants self-identified as female, and about 50% identified as belonging to diverse ethnocultural communities beyond English and French. Most of the participants were nurses, social workers, LTC activity aides, counsellors, and therapists.

At the personal and professional levels, the hero discourse has posed many psychological dilemmas for HCPs. They are expected to be strong, invincible, all-enduring, and willing to sacrifice their personal well-being for the common good. When faced with the challenges of sustained long work hours, workplace chaos, physical and emotional exhaustion, and frequent experiences of vicarious trauma, this glorified expectation of self-sacrifice discourages HCPs from recognizing and embracing their limitations, leading to negative self-judgment, feelings of guilt, and a sense of incompetence that further compromise their mental health. In addition to workplace demands, they also have to attend to the physical, psychosocial, and spiritual needs of their family members and loved ones at home. When the hero discourse creates a spotlight solely on the professional identities and roles of HCPs, little attention is afforded to the multiple interdependent identities, experiences, and mental health needs of HCPs beyond their workplace. The results of this study enable us to highlight the seldom explored research gap of understanding the experiences of HCPs as both service providers and service users.

HCPs living with elderly family members or family members with weakened immune systems often function as protective agents as well as caregivers at home. During pandemic lockdowns, HCPs who are family caregivers of children or adults living with disabilities and special needs have to carry out specialized care, and the boundary between family caregiving and service provision becomes blurred. The required efforts to perform specialized care at home have created a double-bind paradox in which they have become unpaid service providers in their personal lives, while also having to maintain their highly demanding health care provision role at work. The multilevel intersecting stressors had taken a toll on the physical

and mental health of HCPs, resulting in sleep disturbance, bodily pain, anxiety, anger, emotional exhaustion, burnout, and even suicidal ideation. Indeed, the Registered Nurses Association of Ontario (RNAO) Work and Wellness Survey [31] showed that 60% of respondents reported very high to high level of stress.

At the organizational level, the hero discourse has also created additional challenges for HCPs. Expressions of concern about workplace pandemic protocols, staff deployment, and shortage of protective equipment as well as questions about organization transparency may be dismissed as unimportant or a sign of individualistic selfishness. As indicated by some study participants, the hero discourse does not address HCPs' frustration and distrust toward policy makers and institutions. Decades of neoliberal practices of austerity, decentralization, privatization, underfunded regulatory regimes, a culture focused on financial efficiency, and increasing controlled and standardized care work in the health and social care sectors have contributed to this distrust [28,32], which is heightened by the chaos, uncertainty, and unprecedented demands of the pandemic. Furthermore, despite making up over 80% of the health care workforce, there are disproportionately fewer women and racialized HCPs in leadership and decision-making positions within the health care system [33]. These structural inequities may help to explain the distrust and the sense of powerlessness expressed by participants. In addition, although established evidence shows that empowerment work environments and meaningful engagement of staff in decision-making are critical mitigating strategies to reduce burnout among HCPs [34,35], policy makers within the health care system have failed to address the challenge of burnout, leaving HCPs at increased vulnerability to burnout and distress during the pandemic [36]. Indeed, 13% of RNs aged 26 years to 35 years reported in the RNAO Work and Wellness Survey that they were very likely to leave the profession after the pandemic [31].

At the community level, many HCPs also experience moral distress [37] associated with a collective visceral resonance of fear, grief, empathy, stress, and anger as they witnessed the devastating impact of COVID-19 (death, loss, social isolation, xenophobia, and racism) on their patients, service users, co-workers, as well as vulnerable groups in the community. For some HCPs, the experience of moral distress intensified during the pandemic lockdowns when medicating children with behavioral challenges (attention deficit hyperactivity disorder, autism spectrum disorder) seemed to be the only viable option to prevent mental health crises of the affected families. For others, the frequent witnessing of patients "dying alone" and having to comfort them with "the touch of a gloved hand" are poignant reminders that the pandemic is not merely a communicable disease that threatens life and brings death; it is an unprecedented phenomenon that disrupts our taken-for-granted sociocultural values and desired practices of dying with dignity, surrounded by loved ones, and comforted by human connections of touch. The witnessing and partaking in these difficult situations left some participants with a sense of helplessness. The so-called pandemic new normal is anything but normal when services are shaped by the fast-changing protocols across the education, community care, and public health sectors, in which care work is deeply rooted in shared

humanness between the care providers and care recipients [38,39].

At the societal or public policy level, the contributions of HCPs were not valued. During the COVID-19 pandemic, HCPs (especially nurses) have been hailed as "unsung heroes" by politicians, the mass media, and the general public [17]. Although the hero discourse has raised public recognition of the important contributions of HCPs, it has not amounted to anything to improve the well-being and pay equity of HCPs. Since November 2019, HCPs in Ontario had been calling the government to repeal Bill 124 without any success. Bill 124 restricts the wages of HCPs (who are mostly women) to a maximum of 1% even though the inflation rate has been over 2% over a decade, and other frontline professionals such as police and firefighters (who are mostly men) are not subject to this bill [17]. Bill 124 illustrates that the hero discourse is merely an empty rhetoric that does not provide any support needed by HCPs. In the Work and Wellbeing Survey of over 2100 nurses, conducted by the RNAO in early 2021, 81% of the respondents reported receiving very poor to fair support from the government, and 59% reported receiving very poor to fair support from their employers [31].

Amid the challenges and demands at work, some participants expressed their desire and commitment to ensure quality care for their patients and service users. For some, the workplace struggles served as a catalyst in strengthening their interpersonal connections with colleagues. The "we are all in this together" spirit seemed to have provided them with a sense of stability amid all the chaos and struggles. Indeed, many participants expressed a deep sense of altruism or responsibility to provide instrumental, psychological, and emotional support to their peers and those affected by the pandemic. This altruistic desire to help others seemed to help give meaning to their stressful work. Thus, the hero discourse is paradoxical in that it functions both as a stressor and a protective anchor. However, altruistic contributions and sustained stressors are often accompanied by adverse health consequences as indicated in the previous paragraphs.

Implications for Policy and Practice

Consistent with the results of this study, 2 decades of research on large-scale epidemics and pandemics have shown persistent and pervasive negative personal impacts on the mental health of HCPs, including increased anxiety, burnout, and posttraumatic stress symptoms [40-42]. In addition, there is established evidence to show that organizational factors like staffing shortage, high workload demands, high work stress, low teamwork, low supervisor support, and work-life imbalance contribute to burnout and high rates of turnover [43-45]. In the fourth quarter of 2020, the job vacancy rate in health and social care increased to 4.7% [46]. The negative impact of COVID-19 on the mental health of HCPs has likely intensified the job vacancy rate. Thus, during a pandemic, it is imperative for health and social care organizations to attend to the mental health and emotional needs of HCPs.

Since the mental health of HCPs during the pandemic was influenced by complex and dynamic interplay of personal, interpersonal, organizational, community, and societal factors,

a multiprong approach is necessary to mitigate these complex stressors. The RNAO Work and Wellness Survey [31] showed that, while 60% of respondents reported very high to a high level of stress, only 8.6% used mental health support through their employers, and only 1.1% used the provincial helpline. Instead, 49.6% sought support from their spouse, family, and friends, and 28.5% sought support from their colleagues. These results suggest that structured peer support programs at work may be an important strategy to maintain mental health of HCPs during a pandemic. Other organizational support such as flexible scheduling, destigmatization of help-seeking, and free programs that support HCPs to develop self-awareness of stress, self-care, and stress reduction are important. Free programs that apply mindfulness, music therapy, relaxation techniques, online workout, peer support, and one-on-one counselling have been found to be effective in stress reduction [47-49].

Additionally, providing timely and effective multiway communication and feedback can enhance social relationships and foster strong social support networks for HCPs within their organizations. Frequent timely communication of accurate, clear, and accessible information about the pandemic with HCPs of all levels is critical. At the same time, mechanisms for HCPs to report concerns, ask questions, and provide feedback to changing protocols are critical. Organizations may benefit from the real-time insights and feedback from the staff to refine their changing protocols. Although it is often impossible for organizations to adopt all suggestions from all staff, thoughtful responses to questions and concerns can be constructive in building and maintaining a collective vision, especially during a period of uncertainty and rapid change. Using peer liaison leaders in every department can enhance effective communication, build trust, and promote a sense of belonging among staff.

Finally, at the societal or public policy level, pandemic preparedness needs to be underpinned by the principles of social justice and equity. Across Canada, over 80% of the COVID-19–related deaths occurred in LTC homes [50]. Low wages, scarce sick benefits, and part-time employment meant that many LTC staff were working in multiple LTC homes, and some might have continued to work while they were ill [50,51]. Thus, existing inequitable distribution of resources within the health care system (eg, hospital vs LTC facilities), inadequate regulation of LTC homes, pay inequity for nursing staff, and the lack of access to full-time employment experienced by racialized staff in the lower ranks of the HCP hierarchy must be addressed.

Limitations

In this study, we engaged participants who are reflective of the HCP workforce in Canada (ie, 85% self-identified as female), as well as ethnoracially diverse participants reflective of the Greater Toronto Area (close to 50% self-identified as belonging to visible minority groups). We have drawn on the qualitative reflection of participants to provide insights on the multilevel, intersecting sources of stressors experienced by HCPs amid a global pandemic. However, there are some study limitations. Since we used convenience and snowball sampling in recruitment, the range of HCP participants we engaged with

might be limited. Furthermore, the self-reflective method did not allow us to engage in more in-depth exploration of specific phenomena that might have emerged from follow-up questions. As this is a qualitative component of the study, we did not generalize our findings. Rather, we have provided important contexts about the sources of stressors in the different domains of the participants' lives, which are useful to guide follow-up research, programs, and policies in the postpandemic period. To reduce possible biases of the researchers toward a particular perspective, we have presented our theoretical framework, engaged team members with different disciplinary perspectives, and kept an audit trail of our analysis process including transcripts of responses, coding, and notes or recordings documenting decisions made in the study.

In addition, this paper reports only on the results of the qualitative component of a broader intervention study. It focuses on the sources and contexts of stressors experienced by HCPs during the first 2 waves of the pandemic. It does not include results on the feasibility and outcomes of PACER, which are beyond the scope of this paper. However, the evaluation protocols of PACER have been published elsewhere [19]. The evaluation of PACER is currently underway to examine the effectiveness of PACER in reducing the psychological impact of stress, improving mental health outcomes, and promoting resilience but is beyond the scope of this paper.

Conclusion

Insights from this study demonstrate that COVID-19 is not merely a communicable disease but is also experienced as a social and political phenomenon. The reflections of HCP participants based on their lived experiences were situated within a social ecological perspective to provide a broader, more holistic understanding that reveals how personal and interpersonal experiences are shaped by power relations at organizational and societal or policy levels. Applying a socioecological lens allows for multilevel analysis that provides a deeper awareness of pre-existing challenges in the health and social care systems that become magnified during the pandemic. From a public health standpoint, it is critical for government and policy makers to engage HCPs in postpandemic discussion and also respond to pre-existing structural issues that contribute to the mental health stressors of HCPs. Finally, interventions from the individual to public policy levels to address the identified stressors must extend beyond the workplace identities and roles of HCP but to also include multidimensional aspects of their lives and consider home, community, and societal contexts. Thus, areas for future research could explore stressors according to the different types of job among HCPs, as well how stressors are experienced by HCPs from different identities, for example, deeper exploration of the stressors on HCPs from racialized or gendered identities and how they are impacted by these stressors.

Authors' Contributions

CS, AA, and JPHW led the data analysis and drafted the manuscript. MKA and PDG conducted the literature review and contributed to data analysis. ATWL, KF, MKLP, and MV provided critical reviews of the thematic analysis and contributed to revision of multiple drafts of the manuscript. All authors have reviewed and agreed that the final version of the manuscript reflects accuracy and integrity of the study.

Conflicts of Interest

None declared.

References

1. Coronavirus disease (COVID-19). Government of Canada. URL: https://www.canada.ca/en/public-health/services/diseases/coronavirus-disease-covid-19.html [accessed 2022-03-06]
2. StatCan COVID-19: Data to insights for a better Canada. Statistics Canada. URL: https://www150.statcan.gc.ca/n1/pub/45-28-0001/452800012020001-eng.htm [accessed 2022-03-06]
3. Stats in brief: 45-28-0001. Government of Canada. URL: https://www150.statcan.gc.ca/n1/en/catalogue/45280001 [accessed 2022-03-06]
4. Vanhaecht K, Seys D, Bruyneel L, Cox B, Kaesemans G, Cloet M, et al. COVID-19 is having a destructive impact on health-care workers' mental well-being. Int J Qual Health Care 2021 Feb 20;33(1):1-6 [FREE Full text] [doi: 10.1093/intqhc/mzaa158] [Medline: 33270881]
5. Zhang H, Shi Y, Jing P, Zhan P, Fang Y, Wang F. Posttraumatic stress disorder symptoms in healthcare workers after the peak of the COVID-19 outbreak: A survey of a large tertiary care hospital in Wuhan. Psychiatry Res 2020 Dec;294:113541 [FREE Full text] [doi: 10.1016/j.psychres.2020.113541] [Medline: 33128999]

6. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors associated with mental health outcomes among health care workers exposed to coronavirus disease 2019. JAMA Netw Open 2020 Mar 02;3(3):e203976 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3976] [Medline: 32202646]

7. Havaei F, Smith P, Oudyk J, Potter GG. The impact of the COVID-19 pandemic on mental health of nurses in British Columbia, Canada using trends analysis across three time points. Ann Epidemiol 2021 Oct;62:7-12 [FREE Full text] [doi: 10.1016/j.annepidem.2021.05.004] [Medline: 34052436]

8. Wilbiks JMP, Best LA, Law MA, Roach SP. Evaluating the mental health and well-being of Canadian healthcare workers during the COVID-19 outbreak. Healthc Manage Forum 2021 Jul 08;34(4):205-210 [FREE Full text] [doi: 10.1177/08404704211021109] [Medline: 34098760]

9. Crowe S, Howard AF, Vanderspank-Wright B, Gillis P, McLeod F, Penner C, et al. The effect of COVID-19 pandemic on the mental health of Canadian critical care nurses providing patient care during the early phase pandemic: A mixed method study. Intensive Crit Care Nurs 2021 Apr;63:102999 [FREE Full text] [doi: 10.1016/j.iccn.2020.102999] [Medline: 33342649]

10. Carmassi C, Foghi C, Dell'Oste V, Cordone A, Bertelloni CA, Bui E, et al. PTSD symptoms in healthcare workers facing the three coronavirus outbreaks: What can we expect after the COVID-19 pandemic. Psychiatry Res 2020 Oct;292:113312 [FREE Full text] [doi: 10.1016/j.psychres.2020.113312] [Medline: 32717711]

11. Lapum J, Nguyen M, Fredericks S, Lai S, McShane J. "Goodbye … Through a Glass Door": emotional experiences of working in COVID-19 acute care hospital environments. Can J Nurs Res 2021 Mar 20;53(1):5-15 [FREE Full text] [doi: 10.1177/0844562120982420] [Medline: 33342299]

12. Nelson H, Hubbard Murdoch N, Norman K. The role of uncertainty in the experiences of nurses during the Covid-19 pandemic: a phenomenological study. Can J Nurs Res 2021 Jun 04;53(2):124-133. [doi: 10.1177/0844562121992202] [Medline: 33541124]

13. Nyashanu M, Pfende F, Ekpenyong MS. Triggers of mental health problems among frontline healthcare workers during the COVID-19 pandemic in private care homes and domiciliary care agencies: Lived experiences of care workers in the Midlands region, UK. Health Soc Care Community 2022 Feb 26;30(2):e370-e376. [doi: 10.1111/hsc.13204] [Medline: 33107131]

14. Shechter A, Diaz F, Moise N, Anstey DE, Ye S, Agarwal S, et al. Psychological distress, coping behaviors, and preferences for support among New York healthcare workers during the COVID-19 pandemic. Gen Hosp Psychiatry 2020;66:1-8 [FREE Full text] [doi: 10.1016/j.genhosppsych.2020.06.007] [Medline: 32590254]

15. Awano N, Oyama N, Akiyama K, Inomata M, Kuse N, Tone M, et al. Anxiety, depression, and resilience of healthcare workers in Japan during the coronavirus disease 2019 outbreak. Intern Med 2020;59(21):2693-2699 [FREE Full text] [doi: 10.2169/internalmedicine.5694-20] [Medline: 33132305]

16. McCarthy J, Gastmans C. Moral distress: a review of the argument-based nursing ethics literature. Nurs Ethics 2015 Feb 10;22(1):131-152. [doi: 10.1177/0969733014557139] [Medline: 25505098]

17. Mohammed S, Peter E, Killackey T, Maciver J. The "nurse as hero" discourse in the COVID-19 pandemic: A poststructural discourse analysis. Int J Nurs Stud 2021 May;117:103887. [doi: 10.1016/j.ijnurstu.2021.103887] [Medline: 33556905]

18. Eisenstadt M, Stapley E, Deighton J, Wolpert M. Adolescent stressors and their perceived effects on mental well-being: a qualitative study. Ment Health Fam Med 2020;16:992-1001. [doi: 10.13140/RG.2.2.15974.24644]

19. Fung K, Liu JJ, Vahabi M, Li AT, Zurowski M, Wong JP. Pandemic Acceptance and Commitment to Empowerment Response (PACER) training: protocol for the development and rapid-response deployment. JMIR Res Protoc 2021 Dec 06;10(12):e33495 [FREE Full text] [doi: 10.2196/33495] [Medline: 34726602]

20. Hayes SC. Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies. Behavior Therapy 2004;35(4):639-665. [doi: 10.1016/S0005-7894(04)80013-3]

21. CHAMP: Mobilizing people living with HIV and allies to champion HIV prevention and care in ethno-racial communities. CATIE. 2016 Jul 15. URL: https://www.catie.ca/prevention-in-focus/champ-mobilizing-people-living-with-hiv-and-allies-to-champion-hiv-prevention [accessed 2022-03-06]

22. Bronfenbrenner U. The Ecology of Human Development. Cambridge, MA: Harvard University Press; 1979.

23. Golden SD, Earp JAL. Social ecological approaches to individuals and their contexts: twenty years of health education & behavior health promotion interventions. Health Educ Behav 2012 Jun;39(3):364-372. [doi: 10.1177/1090198111418634] [Medline: 22267868]

24. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. Health Educ Q 1988;15(4):351-377. [doi: 10.1177/109019818801500401] [Medline: 3068205]

25. Pavon JM, Fish LJ, Colón-Emeric CS, Hall KS, Morey MC, Pastva AM, et al. Towards "mobility is medicine": Socioecological factors and hospital mobility in older adults. J Am Geriatr Soc 2021 Jul 23;69(7):1846-1855 [FREE Full text] [doi: 10.1111/jgs.17109] [Medline: 33755991]

26. Merriam SB. Qualitative Research: A Guide to Design and Implementation. Hoboken, NJ: Jossey-Bass; 2009.

27. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

28. Baines D. Neoliberalism and the convergence of nonprofit care work in Canada. Competition & Change 2015 May 05;19(3):194-209. [doi: 10.1177/1024529415580258]

29. Stavrevska EB. Feminised work, invisible labour: Against the formal-informal economy dichotomy. In: Routledge Handbook of Feminist Peace Research. Milton Park, Abingdon-on-Thames, Oxfordshire, England, UK: Routledge; 2021.

30. Employment by class of worker, annual (x 1,000). Statistics Canada. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410002701 [accessed 2022-03-06]

31. Work and Well-Being Survey Results. RNAO. 2021 Mar. URL: https://rnao.ca/sites/rnao-ca/files/Nurses_Wellbeing_Survey_Results_-_March_31.pdf [accessed 2022-03-06]

32. Church J, Gerlock A, Smith DL. Neoliberalism and accountability failure in the delivery of services affecting the health of the public. Int J Health Serv 2018 Oct 13;48(4):641-662. [doi: 10.1177/0020731418793106] [Medline: 30213248]

33. Bourgeault I, James Y, Lawford K, Lundine J. Empowering women leaders in health: A gap analysis of the state of knowledge. Canadian Journal of Physician Leadership 2018;5(2):1 [FREE Full text]

34. Boamah SA, Read EA, Spence Laschinger HK. Factors influencing new graduate nurse burnout development, job satisfaction and patient care quality: a time-lagged study. J Adv Nurs 2017 May 21;73(5):1182-1195. [doi: 10.1111/jan.13215] [Medline: 27878844]

35. Kilroy S, Flood PC, Bosak J, Chênevert D. Perceptions of high-involvement work practices, person-organization fit, and burnout: a time-lagged study of health care employees. Hum Resour Manage 2016 Aug 31;56(5):821-835. [doi: 10.1002/hrm.21803]

36. Brophy JT, Keith MM, Hurley M, McArthur JE. Sacrificed: Ontario healthcare workers in the time of COVID-19. New Solut 2021 Feb 11;30(4):267-281 [FREE Full text] [doi: 10.1177/1048291120974358] [Medline: 33174768]

37. Cacchione PZ. Moral distress in the midst of the COVID-19 pandemic. Clin Nurs Res 2020 May 04;29(4):215-216. [doi: 10.1177/1054773820920385] [Medline: 32363981]

38. Gillon R. Restoring humanity in health and social care – Some suggestions. Clinical Ethics 2013 Sep 26;8(4):105-110. [doi: 10.1177/1477750913502618]

39. Salhi BA, Osborne AD. Incarceration and social death — restoring humanity in the clinical encounter. N Engl J Med 2021 Jan 21;384(3):201-203. [doi: 10.1056/nejmp2023874]

40. Li X, Zhou Y, Xu X. Factors associated with the psychological well-being among front-line nurses exposed to COVID-2019 in China: A predictive study. J Nurs Manag 2021 Mar;29(2):240-249. [doi: 10.1111/jonm.13146] [Medline: 32890453]

41. Moradi Y, Baghaei R, Hosseingholipour K, Mollazadeh F. Challenges experienced by ICU nurses throughout the provision of care for COVID-19 patients: A qualitative study. J Nurs Manag 2021 Jul 02;29(5):1159-1168 [FREE Full text] [doi: 10.1111/jonm.13254] [Medline: 33480145]

42. Blekas A, Voitsidis P, Athanasiadou M, Parlapani E, Chatzigeorgiou AF, Skoupra M, et al. COVID-19: PTSD symptoms in Greek health care professionals. Psychol Trauma 2020 Oct;12(7):812-819. [doi: 10.1037/tra0000914] [Medline: 32853011]

43. Fiabane E, Giorgi I, Sguazzin C, Argentero P. Work engagement and occupational stress in nurses and other healthcare workers: the role of organisational and personal factors. J Clin Nurs 2013 Sep 29;22(17-18):2614-2624. [doi: 10.1111/jocn.12084] [Medline: 23551268]

44. Hämmig O. Explaining burnout and the intention to leave the profession among health professionals - a cross-sectional study in a hospital setting in Switzerland. BMC Health Serv Res 2018 Oct 19;18(1):785 [FREE Full text] [doi: 10.1186/s12913-018-3556-1] [Medline: 30340485]

45. Toh SG, Ang E, Devi MK. Systematic review on the relationship between the nursing shortage and job satisfaction, stress and burnout levels among nurses in oncology/haematology settings. Int J Evid Based Healthc 2012 Jun;10(2):126-141. [doi: 10.1111/j.1744-1609.2012.00271.x] [Medline: 22672602]

46. Job vacancies, fourth quarter 2020. Statistics Canada. 2021 Mar 23. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/210323/dq210323b-eng.htm [accessed 2022-03-06]

47. Cherepanov E. Responding to the psychological needs of health workers during pandemic: ten lessons from humanitarian work. Disaster Med Public Health Prep 2020 Sep 10:1-7 [FREE Full text] [doi: 10.1017/dmp.2020.356] [Medline: 32907680]

48. Callus E, Bassola B, Fiolo V, Bertoldo EG, Pagliuca S, Lusignani M. Stress reduction techniques for health care providers dealing with severe coronavirus infections (SARS, MERS, and COVID-19): a rapid review. Front Psychol 2020 Dec 10;11:589698 [FREE Full text] [doi: 10.3389/fpsyg.2020.589698] [Medline: 33362654]

49. de Witte M, Pinho ADS, Stams G, Moonen X, Bos AE, van Hooren S. Music therapy for stress reduction: a systematic review and meta-analysis. Health Psychol Rev 2022 Mar 27;16(1):134-159. [doi: 10.1080/17437199.2020.1846580] [Medline: 33176590]

50. Vilches TN, Nourbakhsh S, Zhang K, Juden-Kelly L, Cipriano LE, Langley JM, et al. Multifaceted strategies for the control of COVID-19 outbreaks in long-term care facilities in Ontario, Canada. medRxiv 2021 Mar 31:1-16 [FREE Full text] [doi: 10.1101/2020.12.04.20244194] [Medline: 33330884]

51. Stall NM, Jones A, Brown KA, Rochon PA, Costa AP. For-profit long-term care homes and the risk of COVID-19 outbreaks and resident deaths. CMAJ 2020 Aug 17;192(33):E946-E955 [FREE Full text] [doi: 10.1503/cmaj.201197] [Medline: 32699006]

Abbreviations

ACE: Acceptance and Commitment to Empowerment
ACT: Acceptance and Commitment Therapy
AMO: Academic Medical Organization
CRCC: Canada Research Coordinating Committee
GEP: Group Empowerment Psychoeducation
HCP: health care providers
LTC: long-term care
MSH-UHN: Mount Sinai Hospital – University Health Network
NFRF: New Frontiers in Research Fund
PACER: Pandemic Acceptance and Commitment to Empowerment Response
PPE: personal protective equipment
RN: registered nurse
RNAO: Registered Nurses Association of Ontario
RPN: registered practical nurse

XSL•FO
RenderX

<u>Original Paper</u>

How Vaccine Ambivalence Can Lead People Who Inject Drugs to Decline COVID-19 Vaccination and Ways This Can Be Addressed: Qualitative Study

Ian David Aronson[1], MA, PhD; Alex S Bennett[1], PhD; Mary-Andrée Ardouin-Guerrier[1], EdD; German Rivera-Castellar[1], MS; Brent Gibson[2], PhD; Samantha Santoscoy[2], BA; Brittney Vargas-Estrella[2], BA

[1]Department of Social and Behavioral Sciences, School of Global Public Health, New York University, New York, NY, United States
[2]New York Harm Reduction Educators, New York, NY, United States

Corresponding Author:
Ian David Aronson, MA, PhD
Department of Social and Behavioral Sciences
School of Global Public Health
New York University
708 Broadway
New York, NY, 10003
United States
Phone: 1 (212) 992 6741
Email: ia14@nyu.edu

Abstract

Background: People who inject drugs are disproportionately impacted by SARS-CoV-2 and COVID-19, yet they do not frequently accept vaccination against SARS-CoV-2 when offered.

Objective: This study aimed to explore why people who inject drugs decline free vaccines against SARS-CoV-2 and how barriers to vaccination can potentially be addressed.

Methods: We conducted semistructured qualitative interviews with 17 unvaccinated adult persons who inject drugs during August and September 2021 at a New York City syringe service program, where approximately three-fourth of participants identified as Latino (55%) or African American (22%). Interviews lasted roughly 20 minutes. The interview guide examined reasons for declining vaccination, participants' understanding of COVID-19 risks, and how messages could be developed to encourage vaccine uptake among people who inject drugs.

Results: Participants acknowledged that they faced increased risk from SARS-CoV-2 owing to their injection drug use but feared that long-term substance use may have weakened their health, making them especially vulnerable to side effects. Fears of possible side effects, compounded by widespread medical mistrust and questions about the overall value of vaccination contributed to marked ambivalence among our sample. The desire to protect children and older family members emerged as key potential facilitators of vaccination.

Conclusions: Community-developed messages are needed in outreach efforts to explain the importance of vaccination, including the far greater dangers of COVID-19 compared to possible unintended side effects. Messages that emphasize vaccines' ability to prevent inadvertently infecting loved ones, may help increase uptake. Community-focused messaging strategies, such as those used to increase HIV and hepatitis C virus testing and overdose prevention among people who inject drugs, may prove similarly effective.

XSL•FO
RenderX

Introduction

People who inject drugs face drastically increased risk of exposure to SARS-CoV-2, the virus that causes COVID-19, and are far more likely to experience severe complications, including death, if they develop COVID-19 [1]. Because COVID-19 principally targets the lungs, it can be especially harmful to individuals who experience respiratory toxicity due to opioids [2,3] (this includes overdose-related respiratory depression) [3] and presents additional risk to polysubstance users who inject opioids and smoke tobacco, cannabis, crack cocaine, or other drugs.

People who inject drugs are frequently homeless [4] or living on streets or in unhygienic settings that severely limit their access to clean water and handwashing [5]. Many also experience crowded, poorly ventilated conditions in shelters and group living settings that make social distancing impossible [5]. The close personal contact required to obtain drugs introduces additional risks of SARS-CoV-2 infection [6], which further increases if people share drugs or equipment. Moreover, to reduce the risk of overdose, People who inject drugs are frequently advised not to inject alone because if something went wrong, no one would be around to help them. Unfortunately, while this prevents overdose deaths, it can cause additional COVID-19 risk.

At the same time, owing to stigma [7,8], lack of access to health care [5], and widespread medical mistrust [9] rooted in a history of discrimination and systemic racism, many people who inject drugs, who are most at risk, are highly unlikely to get vaccinated. Prior research indicates that before the current pandemic, people who inject drugs frequently avoided testing and treatment for hepatitis C virus (HCV) because they feared inappropriate or disrespectful treatment from medical staff [9]. Likewise, although hepatitis B virus (HBV) infections are frequently attributed to injection drug use, people who inject drugs are less likely to vaccinate against HBV compared to those who do not inject drugs [10], and among persons who inject drugs who do vaccinate against HBV, two-thirds of individuals do not complete the 3-dose vaccination series [11]. A recent study shows nearly half of individuals with substance use disorders (SUDs) are unwilling to vaccinate against COVID-19 [12], underscoring the critical need for new messaging and enhanced outreach to this population.

For African American and Latino people who inject drugs, the disproportionate impact of COVID-19 can be especially deadly [1,5]. COVID-19 has been concentrated in African American and Latino communities nationwide [13,14] since the start of the pandemic, and the effects are clearly visible in New York City [15-17], where this study was conducted. In parts of the South Bronx and East Harlem—New York City areas with mostly African American and Latino populations—COVID-19 deaths per 100,000 people reached roughly 10 times those in majority White areas of Manhattan [16]. Nonetheless, in a survey published in late 2020, fewer than half of African American respondents and fewer than two-thirds of Latino respondents indicated that they would receive COVID-19 vaccines if they were available free of charge [18]. In a 2021 study, 60% of White respondents said that they would get vaccinated against COVID-19, but only 36% of African Americans indicated that they would [19]. A 2021 nationwide study of vaccination rates by race, ethnicity, and state found that although numbers are improving among underrepresented groups, relative uptake rates of COVID-19 vaccination among White adults were higher by a median factor of 1.3 times compared to African American adults and by a median of 1.3 times compared to Latino adults as well, highlighting the need to promote vaccine equity [20].

There is a rich history of messaging guided by evidence-based theories of behavior change (eg, Social Cognitive Theory [21] and the Information, Motivation, Behavioral Skills Model [22]). This includes studies examining how messages can be tailored for, and evaluated among, specific groups (eg, underserved populations who attend church [23]; African American and Latino adolescents [24]). In the course of multiple National Institutes of Health–funded studies designed to encourage health behavior change among underserved populations, our team has developed a methodology focused on community members' expertise. We first conduct in-depth qualitative interviews to understand barriers to, and facilitators of, specific health behaviors among members of a selected population, and then iteratively develop technology-based interventions in response. We have used this methodology to increase HIV testing among emergency department patients [25,26], to increase HIV and HCV testing and overdose prevention among people who inject drugs [27], and to encourage people who inject drugs to carry naloxone kits to reverse overdose events [28]. Accordingly, to inform technology-based intervention messaging aimed at encouraging African American and Latino people who inject drugs to vaccinate against SARS-CoV-2, our team conducted qualitative interviews to examine why people who inject drugs in low-income, historically disadvantaged communities in New York City choose not to get vaccinated. This study describes our initial findings.

Methods

Overview

A purposive sample of unvaccinated people who inject drugs (n=17) was recruited in August and September 2021 from individuals receiving services at a community-based syringe service program (SSP) with a participant base of mostly African American and Latino people who inject drugs in New York City. Staff from the SSP recruited participants in a drop-in-center at SSP headquarters. Participants were recruited from the population of people who come to the organization for a host of services on a weekly or daily basis. Eligibility criteria for interviewees included injection drug use within the past 90 days, and not having received a vaccination against COVID-19, both assessed via self-report. Once it was determined they were eligible, they were referred to members of our research team who explained details of the study. To ensure both anonymity and to prevent duplicate responses, we used program participants' existing SSP anonymous identifiers as their study identification. After participants provided verbal informed consent, 1 of 2 project staff members—who are

coauthors of the current paper—conducted interviews on site in private rooms at the SSP. Interviews lasted approximately 15-30 minutes and were audio-recorded and transcribed verbatim by a third-party service. Participants received US $20 cash at the end of the interview as compensation for their time.

Interviews were conducted in English and Spanish by project staff, using an interview guide developed by the principal investigator (PI) with input from the full project team, including members of the partnering organization. Preliminary interview topics were developed from an exhaustive literature review on vaccine hesitancy and COVID-19. Two members of the study team who were experienced qualitative interviewers and researchers conducted semistructured interviews that explored both barriers to and facilitators of COVID-19 vaccination including the following: COVID-19 knowledge, sources of information, perceptions of vaccine risk, vaccine uptake, vaccination settings, stigma, medical mistrust, and policy-related barriers to and facilitators of vaccination. As part of the discussion of potential facilitators, interviewers asked participants to suggest messages our team could share with other people who inject drugs to increase vaccination rates. The study team also collected demographics from study participants, including race and ethnicity, gender, and primary language spoken at home.

Transcripts were analyzed through thematic analysis. Following completion of each audio-recorded interview, the digital file was promptly transcribed by external transcription services: REV.com for English language interviews and Datagain Services for interviews conducted in Spanish. The PI made sure interview files did not contain any identifying data before submitting them for transcription. Coding was conducted using MAXQDA qualitative analysis software.

Our team met weekly to conduct preliminary analysis of the interview transcripts to identify broad thematic categories addressed in each interview and to discuss these identified thematic categories which consisted of both a priori constructs (based on the aims of the study and the interview guide) and emerging themes (that were related to the study aims but not specifically anticipated). An initial code list was developed from the interview guide and project aims. To expand this preliminary code list, a small subset of interviews was jointly analyzed by 3 team members, and codes and coding strategies were then discussed by the larger team.

Three team members, including 2 who also conducted interviews, used an inductive or deductive approach to analyzing the interviews using a combination of a priori and emergent code categories to identify some of the barriers to and facilitators of COVID-19 vaccination [29]. Additional codes were added during discussions with community partners at the SSP. Codes were compared for thematic consistency and discrepancies were processed during regular team meetings to ensure intercoder reliability. Two authors then independently coded the remaining 15 transcripts, frequently meeting with the study team after coding sets of five interviews to discuss and resolve any coding discrepancies by consensus.

To characterize the range of reasons for declining a vaccination and the most salient ones, we calculated the frequency with which themes were endorsed by study participants. This is not to imply that our qualitative data can be generalized to larger populations of people who inject drugs or to vaccine decliners, but to better characterize the range of themes that emerged among a population of people who inject drugs in New York City. As other qualitative studies have suggested, infrequently endorsed themes can be as important for designing public health interventions as frequently endorsed ones [30]. Once we reached a point where no new thematic areas were identified in newly conducted interviews, we determined by consensus that saturation had been reached; at this point, we completed scheduled interviews and stopped recruiting new interviewees. This process resulted in a total of 17 interviews.

Ethical Considerations

All procedures, including the interview guide, were approved by a single governing institutional review board, BRANY (protocol number 21-039-524), and the study participants provided informed consent.

Results

Participant Characteristics

Seventeen people who reported injection drug use in the past 90 days and had not been vaccinated were interviewed for this study. Participants identified as male (n=11, 64.7%) and female (n=6, 35.3%). The majority of participants identified as Hispanic or Latino (n=11, 64.7%), 5 (29.4%) identified as Black non-Hispanic, and 1 (5.9%) identified as White non-Hispanic. A chief goal of our research was to keep interviews brief and focused on why participants declined vaccination and how potential barriers to vaccination could be addressed. An additional important goal was to protect the privacy of respondents. As a result, we did not include questions about participants' backgrounds beyond self-reported race and gender.

Barriers to Vaccination

Participants described multiple reasons for not getting vaccinated, ranging from a fear of side effects (both known and as of yet undiscovered), to misinformation and medical mistrust, as well as questions about whether the vaccines were even needed. Participants also reported understanding they faced increased risk of SARS-CoV-2 exposure because of their injection drug use, but said they feared long-term drug use could have weakened their immunity and overall physical condition, rendering them especially vulnerable to negative vaccine side effects. Details of related findings, accompanied by specific quotes illustrating each emerging theme, appear below.

Side Effects (Endorsed by 17 of 17 Interviewees)

This fear of side effects included not only immediate reactions, but also reports of people who took the vaccine and within days or weeks developed other health problems or died. These fears appear compounded by concerns the vaccine was released only a short time after the first cases of COVID-19 were reported: "six months later, we got a vaccine. How did they do that?" [Interviewee 6, Hispanic man]. Accordingly, a number of participants reported a desire to wait and see if additional negative effects would emerge over time.

I don't want to get it, and then, you know, three months from now, all of a sudden, um, people who are 40 and under are dying because they, they're, something with the vaccine, how it interacts with our, our brain chemicals as we reach the age of 40. … I've been waiting to see. You know? I'd say five years would be a good period of time to get all the ducks in a row, to know if it's worth it or not. [Interviewee 14, non-Hispanic White man]

I'm like wait a minute … there's a chance I'm taking. I know it's a- it's a better thing for me, like just do it and get it out the way and- and hope not to catch the virus. But it's another thing for me to think "Wait a minute, but if I do it what will happen to me?" I'm not sure I might get a side effect, a bad side effect, some- you know, detrimental to my life or whatever. [Interviewee 7, Hispanic man]

I literally seen somebody caught a reaction next to me on the bus, just came off from getting the Johnson & Johnson. They never made it out of the bus without the paramedics. So, um... I don't know. I want to, but I don't want to. I don't know how it's gonna look in the future. I don't know how the future will be. I was gonna... I was just gonna put that to the future, but... I'm pretty sure they're gonna come up with something better, with vaccines. [Interviewee 6, Hispanic man]

As highlighted by the last quote, many participants claimed firsthand knowledge of someone who became ill after vaccination. However, it remains unclear how many of these incidents are attributable to the vaccine, and which might still have happened (eg, a person getting sick on a bus) even if the person had not vaccinated.

Mistrust (Endorsed by 9 of 17 Interviewees)

Interview data indicate that other theories are circulating and spread by social media and television, as well as by close friends and family members, and they have created strong barriers to vaccination.

You see the TV. You have Farrakhan, Minister Farrakhan. "Don't take this vaccination. It killed this one, it killed that one." … "Oh, the government is putting a chip inside you" and so on and so on. [Interviewee 15, Non-Hispanic Black man]

Lotta people don't vaccinate because they say, "Oh," 'scuse my language, "Oh, that's bullshit," you know. 'Scuse me for expression. You know, "That's, that's, it's a lie or it's the fake. It's just the government, you know, trying to, another way of keeping track of you," you know, and stuff like that. [Interviewee 3, Hispanic man]

Barriers Specific to People who Inject Drugs (Endorsed by 5 of 17 Interviewees)

Two additional themes that emerged through our interviews appear specific to people who inject drugs. Some participants indicated that maintaining a relatively healthy lifestyle, combined with steps they take to protect themselves from HIV—such as only using new syringes and not sharing injection equipment—would protect them from COVID-19 just as well as vaccination:

I've been doing good with my own regimen. You know, washing my hands, eating right, so on and so on. I've been doing good so far and I figure I don't need it. [Interviewee 15, Non-Hispanic Black man]

Other participants recognized that the close contact of social interactions related to drug use made them increasingly vulnerable to COVID-19 and therefore required additional safeguards. At the same time, some within our sample questioned whether the toll that years or decades of drug use had taken on their physical health made them even more susceptible to potential vaccine side effects.

Especially as drug users … it attacks us first. Because we come in contact with all different kind of people and they be high and they don't take care they life … They don't take, medicate, nothing … So we are high risk. So therefore as, as us being high risk, we should be at the door, knocking at the door like bang, bang, bang, let me in. [Interviewee 13, Non-Hispanic Black woman]

Well I've been a drug user for a lot of years, number of years, and I don't know if my body will be able to take- I'm not- I'm not sure about catching certain side effects that can be detrimental to my life. You know? … I'm scared of that fact that I don't know what to expect. I mean I know I have a better chance at surviving from getting at if I am vaccinated. However … if I do take the vaccination I'm not sure what might happen to me. You know? I'm not sure if I might get a side effect … or die. Die from it because of my drug use over the years. [Interviewee 7, Hispanic man]

Facilitators

Interviews with our sample of unvaccinated people who inject drugs also established a number of potential vaccination facilitators. Participants spoke of wanting to protect family members from inadvertent exposure, especially children who were too young to receive vaccination and older adults who, in general, faced greater risk from COVID-19. Participants also reported that simply discussing the increased risk of COVID-19 exposure among substance users might facilitate vaccination uptake, if handled correctly. Interview data underscore the value of phrasing intervention messages in positive terms (eg, "getting vaccinated will help you because…"), and the importance of not making intervention recipients felt as though vaccination is being forced upon them.

Protecting Family and Members of One's Social Network (Endorsed by 5 of 17 Interviewees)

Children, mothers, and grandmothers all emerged as potential facilitators of vaccination among our sample. As noted earlier, participants expressed a strong desire to protect their families. In addition, participants noted that mothers could prove to be excellent messengers for and models of vaccination uptake.

I would tell them to do it, that they should think about their children, they wouldn't want that to happen to

their children, or anybody else. [Interviewee 1, Hispanic man]

If my mother tells me. She has told me so many times too. She is vaccinated too. My mom. Thank God. And she is fine. … And my grandma. Thank God, they are already fine, and they are vaccinated. [Interviewee 16, Hispanic man]

Importance of Vaccination for People who Inject Drugs (Endorsed by 9 of 17 Interviewees)

Similar to concerns about vaccination among people who inject drugs, which emerged among our sample, themes about messages that would facilitate vaccination among people who inject drugs emerged as well. These included messaging about the increased dangers of injection drug use during the pandemic, which again underscore the risks stemming from the close personal contact associated with injection drug use:

> *I recommend vaccination to any person using drugs on the streets. They should be vaccinated, because they are on the streets, and they don't know who they may infect if they are infected. You don't know … When you use drugs, you interact with many people, and you may infect many people. So, I would tell them to get vaccinated as soon as possible. That's what I say.* [Interviewee 17 Hispanic man]

Importance of Positive Message Tone (Endorsed by 17 of 17 Interviewees)

Suggested methodologies to increase vaccination among people who inject drugs also included simple reassurance as a facilitator, including telling people that they would not be left alone in their time of need.

> *Something like that, "If you're afraid, I'll go with you." Yeah … just to encourage them by saying that it does- I, I, it doesn't hurt, you know. Only take a couple of seconds, and, and it's over. And you're good.* [Interviewee 9, Hispanic man]

Lastly, although our sample only included people who had not vaccinated against COVID-19, some participants indicated that the process of discussing their resistance to vaccination during study interviews and describing how they might explain the importance of vaccination to someone else left them more likely to accept a vaccination if one were offered.

> *I would say, "Listen … with me, myself, I never got the vaccine." I will try 'cause, you know, really, I, I'm scared, but ... I'm sitting here with you, so that's why I'm try- I would like to get the vaccine.* [Interviewee 11, Hispanic woman]

Discussion

Principal Findings

In many ways, our sample may be described as more "vaccine ambivalent" than "vaccine hesitant" or "vaccine resistant." Interview data show participants are clearly aware of their increased risk of COVID-19 exposure, and that participants understand the importance of vaccination to prevent illness. At the same time, participants expressed concerns about the vaccines' safety, and whether a possible weakening of their bodies owing to injection drug use has intensified the danger of possible vaccine side effects. In addition, participants in our sample, like others in our society, describe a seemingly endless stream of social media misinformation, often repeated by close friends and family members, which actively discourages vaccination. Together, these issues have left our participants unsure what steps they should take to protect themselves and what sources of information they can trust.

This ambivalence becomes more important to address as the danger of COVID-19, specifically among substance users, is increasingly well documented. According to a large, nationwide study conducted by the National Institutes of Health and researchers at Case Western Reserve University [31], risk of hospitalization and death from COVID-19 is strongest among those recently diagnosed with opioid use disorder (OUD), which includes heroin injectors. Further, among individuals recently diagnosed with OUD, risks are significantly stronger among African American people who use opioids than among White people who use opioids (risk among Hispanic or Latino people who use opioids was not specifically analyzed by the study) [31]. Thus, given the composition of our sample in which the majority of participants were by far African American or Latino, it would be logical to expect that an understanding of these risks would in itself encourage vaccination. Nonetheless, interview data indicate that barriers to vaccination described above, and in particular those related to potential side effects, strongly discouraged participants from getting vaccinating even when they clearly understood their increased risk from COVID-19 and expressed a desire to protect themselves and people close to them.

Participants repeatedly mentioned that they feared their history of drug use would exacerbate the dangers of potential side effects, possibly to the point that vaccination might prove deadly. These concerns may be particularly difficult to dispel among people who inject drugs, who are contemplating vaccination and, in fact, appear to have scientific merit. Clinical trials for the approved COVID-19 vaccines did not explicitly include participants with SUD, and there are no current systematic studies examining the real-world effectiveness of COVID-19 vaccines among SUD or populations of people who inject drugs [32]. Additionally, a large nationwide analysis published in October 2021 indicates that fully vaccinated people who were diagnosed with SUD are at significantly higher risk for breakthrough infection than fully vaccinated people who have not been diagnosed with SUD [32]. Thus, although it is completely understandable when an interviewee says they want to wait 5 years to see if vaccination is worthwhile and safe, curtailing the spread of COVID-19 among people who use drugs and others at the highest risk of infection requires far more immediate action.

Some interviews suggest participants became more amenable to vaccination during their discussions with our team. This may indicate that simply allowing interviewees to voice their concerns about vaccination without being judged or silenced encouraged participants to consider the benefits of vaccination and left them feeling empowered to make positive decisions

about protecting themselves. A similarly supportive and noncoercive approach could potentially encourage vaccination—and health care utilization in general—on a wider scale among people who inject drugs and other members of marginalized, high-risk populations, especially if paired with messaging designed to address known barriers to vaccination and emphasize facilitators (eg, "No matter what you do we want you to be safe, and the best way to keep yourself, your kids, and people you care about safe from COVID is to vaccinate"). This type of strategy, rooted in the idea of low-threshold access to health care via harm reduction services [33,34], could prove particularly valuable if vaccine mandates are shown ineffective among our target population. Not all the people who inject drugs in our sample have jobs; hence, employer mandates might not fully increase vaccination rates. Similarly, many people in our sample might not seek to dine at restaurants or visit museums; hence, related vaccination requirements may similarly have little effect.

Next Steps

Building upon our team's participatory methodology, we now plan to create a series of digitally delivered intervention messages based on the aforementioned findings, including short videos and SMS text messages to encourage vaccination among people who inject drugs, who initially decline. All steps of intervention development and evaluation will include continued involvement among people who inject drugs, including ongoing monthly meetings with a community advisory board who we have consulted since August 2021.

Limitations

This study is not without limitations, the first one related to the generalizability of a fairly small sample recruited at a New York City SSP. Owing to the relative saturation of area harm reduction services, participants may feel less stigma and may therefore be more open to discussing substance use and related health issues, including decisions to decline vaccination. This may also be a strength of our study, as interviews conducted in a supportive SSP environment may have resulted in more detailed,

and possibly more honest, participant responses. Additionally, our analyses show the themes discussed in this study emerged among a number of interviews. This suggests current data may represent larger topics of concern among people who inject drugs in the New York City area. Further research is warranted to examine whether similar themes would be present among additional populations of people who inject drugs nationwide.

Conclusions

Effectively encouraging vaccination among people who inject drugs remains a public health priority. If members of populations of people who inject drugs, who are generally already greatly underserved by health care, do not get vaccinated in adequate numbers, SARS-CoV-2 can continue to spread and mutate into potentially more dangerous variants. Appeals to protect the health of loved ones and the safety of the larger community have been successfully employed to encourage important health behaviors among populations of people who inject drugs, such as HIV or HCV testing and prevention along with overdose prevention and reversal. Developing intervention materials that depict trusted community members (eg, recognizable peer educators and individuals with compelling personal stories) delivering clear, nonjudgmental messages may prove especially helpful now. Similarly, organizations that currently offer services to people who inject drugs, including SSPs or other community-based organizations, can facilitate vaccination by offering easy access to vaccines on site (people may be more likely to accept a vaccine offer from someone they already know and trust, especially if they do not have to travel to an additional location).

Importantly, data from this study show that people who inject drugs often remain reluctant to get vaccinated even when they understand the dangers posed by COVID-19, and the increased safety vaccines can offer to them and their families. Given historic vaccine hesitance and medical mistrust among people who inject drugs, combined with pervasive concern about possible side effects, interventions are urgently needed to effectively offer vaccines in ways those most in need will actually accept them.

Conflicts of Interest

None declared.

References

1. Iversen J, Sabin K, Chang J, Morgan Thomas R, Prestage G, Strathdee SA, et al. COVID-19, HIV and key populations: cross-cutting issues and the need for population-specific responses. J Int AIDS Soc 2020 Oct;23(10):e25632 [FREE Full text] [doi: 10.1002/jia2.25632] [Medline: 33119183]
2. Schimmel J, Manini AF. Opioid Use Disorder and COVID-19: Biological Plausibility for Worsened Outcomes. Subst Use Misuse 2020;55(11):1900-1901 [FREE Full text] [doi: 10.1080/10826084.2020.1791184] [Medline: 32657207]

XSL•FO
RenderX

3. Volkow ND. Collision of the COVID-19 and Addiction Epidemics. Ann Intern Med 2020 Jul 07;173(1):61-62 [FREE Full text] [doi: 10.7326/M20-1212] [Medline: 32240293]

4. Doran KM, Rahai N, McCormack RP, Milian J, Shelley D, Rotrosen J, et al. Substance use and homelessness among emergency department patients. Drug Alcohol Depend 2018 Jul 01;188:328-333 [FREE Full text] [doi: 10.1016/j.drugalcdep.2018.04.021] [Medline: 29852450]

5. Vasylyeva TI, Smyrnov P, Strathdee S, Friedman SR. Challenges posed by COVID-19 to people who inject drugs and lessons from other outbreaks. J Int AIDS Soc 2020 Jul;23(7):e25583 [FREE Full text] [doi: 10.1002/jia2.25583] [Medline: 32697423]

6. Abadie R, Gelpi-Acosta C, Aquino-Ruiz F, Aponte-Melendez Y. COVID-19 risks among people who inject drugs in Puerto Rico. Int J Drug Policy 2021 Jul;93:102903 [FREE Full text] [doi: 10.1016/j.drugpo.2020.102903] [Medline: 33069554]

7. van Boekel LC, Brouwers EP, van Weeghel J, Garretsen HF. Stigma among health professionals towards patients with substance use disorders and its consequences for healthcare delivery: systematic review. Drug Alcohol Depend 2013 Jul 01;131(1-2):23-35. [doi: 10.1016/j.drugalcdep.2013.02.018] [Medline: 23490450]

8. Barocas JA, Brennan MB, Hull SJ, Stokes S, Fangman JJ, Westergaard RP. Barriers and facilitators of hepatitis C screening among people who inject drugs: a multi-city, mixed-methods study. Harm Reduct J 2014 Jan 14;11:1 [FREE Full text] [doi: 10.1186/1477-7517-11-1] [Medline: 24422784]

9. Strauss SM, Munoz-Plaza C, Tiburcio NJ, Astone-Twerell J, Des Jarlais DC, Gwadz M, et al. Barriers and Facilitators to Undergoing Hepatitis C Virus (HCV) Testing through Drug Treatment Programs. J Drug Issues 2008 Oct 01;38(4):1161-1185. [doi: 10.1177/002204260803800411]

10. Harris AM, Iqbal K, Schillie S, Britton J, Kainer MA, Tressler S, et al. Increases in Acute Hepatitis B Virus Infections - Kentucky, Tennessee, and West Virginia, 2006-2013. MMWR Morb Mortal Wkly Rep 2016 Jan 29;65(3):47-50 [FREE Full text] [doi: 10.15585/mmwr.mm6503a2] [Medline: 26821369]

11. Tressler S, Bhandari R. Interventions to Increase Completion of Hepatitis B Vaccination in People who Inject Drugs: A Systematic Review and Meta-analysis. Open Forum Infect Dis 2019 Dec;6(12):ofz521 [FREE Full text] [doi: 10.1093/ofid/ofz521] [Medline: 31890724]

12. Mellis AM, Kelly BC, Potenza MN, Hulsey JN. Trust in a COVID-19 vaccine among people with substance use disorders. Drug Alcohol Depend 2021 Mar 01;220:108519 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.108519] [Medline: 33461150]

13. Millett GA, Honermann B, Jones A, Lankiewicz E, Sherwood J, Blumenthal S, et al. White Counties Stand Apart: The Primacy of Residential Segregation in COVID-19 and HIV Diagnoses. AIDS Patient Care STDS 2020 Oct;34(10):417-424 [FREE Full text] [doi: 10.1089/apc.2020.0155] [Medline: 32833494]

14. Adhikari S, Pantaleo NP, Feldman JM, Ogedegbe O, Thorpe L, Troxel AB. Assessment of Community-Level Disparities in Coronavirus Disease 2019 (COVID-19) Infections and Deaths in Large US Metropolitan Areas. JAMA Netw Open 2020 Jul 01;3(7):e2016938 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.16938] [Medline: 32721027]

15. Ross J, Diaz CM, Starrels JL. The Disproportionate Burden of COVID-19 for Immigrants in the Bronx, New York. JAMA Intern Med 2020 Aug 01;180(8):1043-1044. [doi: 10.1001/jamainternmed.2020.2131] [Medline: 32383754]

16. COVID-19: Data. City of New York. URL: https://www1.nyc.gov/site/doh/covid/covid-19-data-totals.page [accessed 2020-12-23]

17. Ogedegbe G, Ravenell J, Adhikari S, Butler M, Cook T, Francois F, et al. Assessment of Racial/Ethnic Disparities in Hospitalization and Mortality in Patients With COVID-19 in New York City. JAMA Netw Open 2020 Dec 01;3(12):e2026881 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.26881] [Medline: 33275153]

18. Coronavirus Vaccine Hesitancy in Black and Latinx Communities. Candid. 2020 Nov 27. URL: https://philanthropynewsdigest.org/features/research-briefs/coronavirus-vaccine-hesitancy-in-black-and-latinx-communities [accessed 2022-03-11]

19. Lee T, Chen AH. Last-Mile Logistics of Covid Vaccination - The Role of Health Care Organizations. N Engl J Med 2021 Feb 25;384(8):685-687. [doi: 10.1056/NEJMp2100574] [Medline: 33503340]

20. Reitsma MB, Goldhaber-Fiebert JD, Salomon JA. Quantifying and Benchmarking Disparities in COVID-19 Vaccination Rates by Race and Ethnicity. JAMA Netw Open 2021 Oct 01;4(10):e2130343 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.30343] [Medline: 34668949]

21. Locke EA. Social Foundations of Thought and Action: A Social-Cognitive ViewSocial Foundations of Thought and Action: A Social-Cognitive View, by Bandura Albert. Englewood Cliffs, NJ: Prentice-Hall, 1986, 617 pp., cloth. AMR 1987 Jan;12(1):169-171. [doi: 10.5465/amr.1987.4306538]

22. Fisher JD, Fisher WA. Theoretical Approaches to Individual-Level Change in HIV Risk Behavior. In: Handbook of HIV Prevention. Boston, MA: Springer; 2000:3-55.

23. Campbell MK, Hudson MA, Resnicow K, Blakeney N, Paxton A, Baskin M. Church-based health promotion interventions: evidence and lessons learned. Annu Rev Public Health 2007;28:213-234. [doi: 10.1146/annurev.publhealth.28.021406.144016] [Medline: 17155879]

24.   Jemmott JB, Jemmott LS, Braverman PK, Fong GT. HIV/STD risk reduction interventions for African American and Latino adolescent girls at an adolescent medicine clinic: a randomized controlled trial. Arch Pediatr Adolesc Med 2005 May;159(5):440-449. [doi: 10.1001/archpedi.159.5.440] [Medline: 15867118]

25.   Aronson ID, Zhang J, Rajan S, Marsch LA, Bugaighis M, Ibitoye MO, et al. Automated Substance Use/Sexual Risk Reporting and HIV Test Acceptance Among Emergency Department Patients Aged 13-24 Years. AIDS Behav 2021 Oct 27. [doi: 10.1007/s10461-021-03507-2] [Medline: 34705152]

26.   Aronson ID, Marsch LA, Rajan S, Koken J, Bania TC. Computer-based video to increase HIV testing among emergency department patients who decline. AIDS Behav 2015 Mar;19(3):516-522 [FREE Full text] [doi: 10.1007/s10461-014-0853-5] [Medline: 25064047]

27.   Aronson ID, Bennett A, Marsch LA, Bania TC. Mobile Technology to Increase HIV/HCV Testing and Overdose Prevention/Response among People Who Inject Drugs. Front Public Health 2017;5:217 [FREE Full text] [doi: 10.3389/fpubh.2017.00217] [Medline: 28879174]

28.   Bennett AS, Freeman R, Des Jarlais DC, Aronson ID. Reasons People Who Use Opioids Do Not Accept or Carry No-Cost Naloxone: Qualitative Interview Study. JMIR Form Res 2020 Dec 23;4(12):e22411 [FREE Full text] [doi: 10.2196/22411] [Medline: 33355094]

29.   Engle M. Qualitative Data Analysis: An expanded Sourcebook (2nd Ed.) Matthew B. Miles and A. Michael Huberman. Thousand Oaks, CA: Sage publications, 1994, 336 pp. Am J Eval 1999;20(1):159-160. [doi: 10.1016/s1098-2140(99)80125-8]

30.   Duhart Clarke SE, Kral AH, Zibbell JE. Consuming illicit opioids during a drug overdose epidemic: Illicit fentanyls, drug discernment, and the radical transformation of the illicit opioid market. Int J Drug Policy 2022 Jan;99:103467 [FREE Full text] [doi: 10.1016/j.drugpo.2021.103467] [Medline: 34662847]

31.   Wang QQ, Kaelber DC, Xu R, Volkow ND. COVID-19 risk and outcomes in patients with substance use disorders: analyses from electronic health records in the United States. Mol Psychiatry 2021 Jan;26(1):30-39 [FREE Full text] [doi: 10.1038/s41380-020-00880-7] [Medline: 32929211]

32.   Wang L, Wang Q, Davis PB, Volkow ND, Xu R. Increased risk for COVID-19 breakthrough infection in fully vaccinated patients with substance use disorders in the United States between December 2020 and August 2021. World Psychiatry 2022 Feb;21(1):124-132 [FREE Full text] [doi: 10.1002/wps.20921] [Medline: 34612005]

33.   Heller D, McCoy K, Cunningham C. An invisible barrier to integrating HIV primary care with harm reduction services: philosophical clashes between the harm reduction and medical models. Public Health Rep 2004;119(1):32-39 [FREE Full text] [doi: 10.1177/003335490411900109] [Medline: 15147647]

34.   Fox AD, Chamberlain A, Frost T, Cunningham CO. Harm Reduction Agencies as a Potential Site for Buprenorphine Treatment. Subst Abus 2015;36(2):155-160 [FREE Full text] [doi: 10.1080/08897077.2015.1011820] [Medline: 25837290]

Abbreviations

HBV: hepatitis B virus
HCV: hepatitis C virus
OUD: opioid use disorder
PI: principal investigator
SSP: syringe service program
SUD: substance use disorders

XSL•FO

RenderX

<u>Original Paper</u>

Competition and Integration of US Health Systems in the Post-COVID-19 New Normal: Cross-sectional Survey

Jiban Khuntia[1], PhD; Xue Ning[2], PhD; Rulon Stacey[1], PhD

[1]Business School, University of Colorado Denver, Denver, CO, United States

[2]Business Department, University of Wisconsin Parkside, Kenosha, WI, United States

Corresponding Author:
Jiban Khuntia, PhD
Business School
University of Colorado Denver
1475 Lawrence St.
Denver, CO, 80202
United States
Phone: 1 303 315 8424
Email: jiban.khuntia@ucdenver.edu

Abstract

Background: How do health systems in the United States view the concept of merger and acquisition (M&A) in a post-COVID 19 "new normal"? How do new entrants to the market and incumbents influence horizontal and vertical integration of health systems? Traditionally, it has been argued that M&A activity is designed to reduce inequities in the market, shift toward value-based care, or enhance the number and quality of health care offerings in a given market. However, the recent history of M&A activity has yielded fewer noble results. As might be expected, the smaller the geographical region in which M&A activity is pursued, the higher the likelihood that monopolistic tendencies will result.

Objective: We focused on three types of competition perceptions, external environment uncertainty–related competition, technology disruption–driven competition, and customer service–driven competition, and two integration plans, vertical integration and horizontal integration. We examined (1) how health system characteristics help discern competition perceptions and integration decisions, and (2) how environment-, technology-, and service-driven competition aspects influence vertical and horizontal integration among US health systems in the post-COVID-19 new normal.

Methods: We used data for this study collected through a consultant from a robust group of health system chief executive officers (CEOs) across the United States from February to March 2021. Among the 625 CEOs, 135 (21.6%) responded to our survey. We considered competition and integration aspects from the literature and ratified them via expert consensus. We collected secondary data from the Agency for Healthcare Research and Quality (AHRQ) Compendium of the US Health Systems, leading to a matched data set for 124 health systems. We used inferential statistical comparisons to assess differences across health systems regarding competition and integration, and we used ordered logit estimations to relate competition and integration.

Results: Health systems generally have a high level of the four types of competition perceptions, with the greatest concern being technology disruption–driven competition rather than environment uncertainty–related competition and customer service–driven competition. The first set of estimation results showed that size, teaching status, revenue, and uncompensated care burden are the main contingent factors influencing the three competition perceptions. The second set of estimation results revealed the relationships between different competition perceptions and integration plans. For vertical integration, environment uncertainty–related competition had a significant positive influence ($P<.001$), while the influence of technology disruption–driven competition was significant but negative ($P<.001$). The influence of customer service–driven competition on vertical integration was not evident. For horizontal integration, the results were similar for environment uncertainty–related competition and technology disruption–driven competition; however, the significance of technology disruption–driven competition was weak ($P=.05$). The influence of customer service–driven competition in the combined model was significant and negative ($P<.001$).

Conclusions: Competition-driven integration has subtle influences across health systems. Environment uncertainty–related competition is a significant factor, with underlying contingent factors such as revenue concerns and leadership as the leading causes of integration plans. However, technology disruption may hinder integrations. Undoubtedly, small- and low-revenue health systems facing a high level of competition are likely to merge to navigate the health care business successfully. This trend should be a focus of policy to avoid monopolistic markets.

XSL·FO
RenderX

Introduction

Background

The COVID-19 pandemic in 2020 transformed several aspects of the health care industry. Across the United States, health systems had to activate emergency plans, and cancel elective procedures, patient visits, and many nonessential activities, all while adopting remote and virtual communication care delivery models. As a result of the pandemic, providers have faced many new challenges and opportunities. Some of the financial and operational challenges have led to integrations among health systems to survive in the postpandemic "new normal," with several health systems planning mergers and acquisitions (M&A) involving billions of dollars.

The rise in health care M&A is not entirely new; indeed, it has been increasing over the last decade, with total deals amounting to US $200 billion [1,2]. In 2020-2021 alone, several mergers worth billions of dollars have been in the spotlight. Recent announcements of UnitedHealth's US $13 billion acquisition of Change Healthcare, Centene's US $2.2 billion purchase of Magellan Health, Anthem's deal for MMM Holdings in Puerto Rico, and Cigna's acquisition of the urgent care telehealth provider MDLIVE exemplified both horizontal and vertical integration progressions in the US health care industry in 2021.

M&A reflects two underlying phenomena in the health care business: (1) competitive dynamics and struggle for survival, and (2) integration to solve competitive threats. This study focuses on relating competition perceptions and health systems' integration plans, using data reported by chief executive officers (CEOs) in early 2021.

The objective of this study was two-fold. First, we sought to examine how health system characteristics lead to competition perceptions among health systems, as reported by CEOs in 2021. We assessed the differences between environment-, technology-, and service-driven competition perceptions in health systems with different characteristics, including size, region, ownership status, teaching status, revenue, number of physicians, and number of hospitals, among other factors.

The second objective of this study was to examine how these three types of competition perceptions influence vertical and horizontal integrations of US health systems in the post-COVID-19 new normal. Delineating competition perceptions and integration plans will help guide strategies and policies in health care.

Competition Perceptions and Integration Plans

Beyond the recent disruptions due to COVID-19, health systems have been facing at least three types of competitive threats over the last decade, driving integrations more than ever before. First, uncertainties stemming from scientific developments have kept some health systems at the forefront of treatment, while others

follow the developments [3]. For example, Detroit-based Henry Ford Health System and East Lansing–based Michigan State University partnered (as an integration, but not a merger) in 2020 to establish a fully integrated cancer program [4]. Similarly, fueled by expanding scientific horizons to care delivery, Atrium Health and Wake Forest Baptist Health merged to form a next-generation academic health system [5].

Second, technological developments have led to new solutions, which have led to new startups in the health care space, such as recently emerging remote and virtual care delivery firms and technology-enabled homecare delivery models (eg, DispatchHealth). The mergers of GigCapital2, UpHealth, and Cloudbreak Health as a unified telemedicine solution provider, and Teladoc's acquisitions of several other technology-enabled models such as Livongo and InTouch Health, exemplify the competition and subsequent mergers due to technological imperatives [6].

Third, capturing a more significant share of patients' care choices across the disease and life continuum increases revenue prospects. In other words, patient–service scope and scale have emerged to fuel competition among health systems. Salt Lake City–based Cimarron Healthcare's acquisition of Ascent Behavioral Health Services, Monroe Capital, and Veronis Suhler Stevenson to expand into therapy and wellness areas is an example [7]. Similarly, the joint venture between Kindred Healthcare of Kentucky and Landmark Medical Center in Rhode Island (a subsidiary of Prime Healthcare Services from California) to own and operate a rehabilitation hospital showcases service expansion competition and integration imperatives [8].

Thus, attempts to adapt and change with innovation, while resisting specialized new generations of competitors on the one hand and the legacy burden of high-cost conglomerate structures to provide everything from high-level intensive care units to quick clinics colocated in drugstores as patient services on the other hand, have aggravated the competitive landscape for health systems [9]. Several very large health systems can bind or lock patients into their ecosystem, with a goal to emerge as a one-stop shop for the entirety of a patient's health care needs and encounters "from the cradle to the grave," in turn encouraging patients to think of the health system for all their health care needs [10]. Some of the examples we provide also reflect that competition has become more asymmetric and is expanding to organizations that are not in health care, such as technology, supply chain, or logistics firms [11].

Ignoring competition looming on the horizon will perish any health system, although managing the problem is not easy. Structural, operational, and strategic issues have only compounded the concerns. Health systems cannot move away from being all things to consumers. Competing in the health care business while lowering costs and improving efficiency

may be possible, but only if health system leaders are innovative and proactive in integrating both high- and low-acuity care options into an effective strategy that seeks to capture a more significant share of patients' health care service continuums and drive subsequent revenue opportunities. Two broad strategies help facilitate these goals: *vertical and horizontal integrations*.

Vertical integration expands the business by acquiring another company that operates before or after the acquiring company in the value chain. In contrast, horizontal integration is when a business grows by acquiring a similar company in their industry at the same value chain point [12]. Often, vertical integration is also done to increase network size and geographic coverage to mitigate risk in contracting, and to achieve market power over buyers and suppliers [13]. Small health systems or physician groups, which align with nonphysician partners such as hospitals, universities, medical schools, and health plans in health care, are touted as vertical integrations. In contrast, merging large or small health systems with similar expertise is an example of horizontal integration [14].

Horizontal integration might lead to enhanced operating efficiencies and economies of scale, improved coordination and quality of patient care, more in-house all-inclusive services, and higher-revenue services (eg, outpatient surgeries, imaging services) [15]. Through megamergers between large health systems, horizontal integration can thwart leaner niche competitors. Integration is challenging in any form because monitoring, coordination, and creating a cohesive culture across the integrated entities are always tricky. In addition, scope economies either do not exist or worsen the integrated organization's dynamics.

The US health care industry has been characterized by different aspects of horizontal and vertical M&A. Such consolidations have changed the structure of US health care markets over time. Care integration has been an emerging solution to improve health quality and reduce costs, especially for a growing population of chronically ill patients in the United States [16]. More specifically, horizontal consolidation occurs when hospitals or physician groups merge, enabling the combined entity to increase its market share. Mergers reduce competing hospitals in a market, and vertical consolidations through acquisition of physician practices by a hospital reduce competition by reducing physicians vying for patients [17].

More structurally integrated organizations may manage care, coordinate across the care continuum, and exploit economies of scale and scope; such capabilities should, in theory, improve health care quality and lower costs. Practitioners and policymakers have therefore made investments in horizontal and vertical integrations. Vertical integration has been associated with better quality and is often framed as optimal care for specific conditions. Stakeholders perceive the positive impacts of consolidation on the long-term viability of health care facilities and their ability to adopt new care models, enhanced competition in health insurance, the creation of foundations, and pioneering medical research and innovation [18]. Stakeholders also believe that consolidation has changed geographic access to care; how physicians make referrals; and

how educated patients are about care, the advertising environment, and economies in surrounding neighborhoods [19]. Market concentration also provides some benefits such as influencing utilization or readmissions and other potential health benefits [20].

Approaches to improve health system performance have been implemented to address care coordination problems and physician burnout. A widely advocated solution is the development of more integrated health delivery systems. While uncertainty remains about the drivers of organizational changes, these changes have led to health care organizations becoming more extensive and more financially integrated [21]. Horizontal and vertical integrations of hospitals and physicians have occurred rapidly for more than two decades. The COVID-19 pandemic may have exacerbated integration plans. Financial sustainability challenges because of declines in utilization and revenue emerged during COVID-19; the economic incentives for hospitals to merge, acquire physician practices, and employ more physicians are increasing. The financial and operational challenges faced during COVID-19 may incentivize health care organizations to make timely decisions for survival. Health systems might gain more significant market power through integration, resulting in greater competitive advantages [17].

Theoretical Framework on Competition and Integration

Transaction cost economics theory has been applied in the health care context, and can plausibly explain the nuances associated with competition and integration dynamics [22]. Anchoring to the transaction costs perspective, this study proposes hazards exchange, transaction efficiency, and cost leadership as the three mechanisms behind the pathway from competition perceptions to integration plans. First, prior literature reveals that the transaction cost perspectives mostly have a focus on exchange hazards and can be used to explain the contracting between hospital groups [22], which forms the competition perceptions of a health system against another one, especially in uncertain environments. Second, transaction cost perspectives consider the efforts and costs required to complete the exchange and contracting activities [22]. Advanced information technologies such as cloud computing can significantly change the operation efficiency and cost, thus changing the economic paradigm and the relationship between collaborative parties [23]. This leads to the consideration of technology disruption–driven competition. Third, as health care is a service-based industry, the application of transaction cost perspectives in the health care context should consider the factor of customer service when perceiving the relationship with partners [24]. The competitive advantage framework proposed in extant research [25] considers three generic strategies: low-cost leadership, differentiation, and focus. This framework reveals the connection between transaction cost economics principles and competition categories, and highlights the significance of the target market or customer to sustain superior performance. This is consistent with customer service–driven competition. In other words, the needs of exchanging hazard in the external environment uncertainty–relevant competition results in integration, the efforts of improving transaction efficiency in the technology disruption–driven competition lead to integration, and the

strategy of cost leadership in the customer service–driven competition causes integration.

More specifically, from the transaction cost economics perspective, vertical integration is about the economies of scope and horizontal integration is about the economies of scale [15]. In order to respond to the potential competition risks and change the competitive status, health systems are seeking either hospital-hospital mergers or hospital-physician integration [26]. Through the acquisitions of physician groups, health systems can expand the scientific horizons to care delivery [5]. Through the mergers of hospitals with similar focus, physicians in health systems can collaborate with similar expertise on a specific program [4]. Both integrations can reduce the hazards caused by uncertain environment–related competition. With higher perception, health systems will seek more unified solutions through vertical integration [6] and change the care delivery model through horizontal integration (eg, DisPatchHealth). The underlying mechanism for both cases is the improvement of transaction efficiency. However, such an outcome may cause even higher perception of technology-driven competition. Prior research suggests a positive relationship between CEO competition perceptions and integration, which arguably needs to be revisited given the recent post-COVID-19 "normal" contexts addressed in this study [27]. The perceptions of competition by CEOs of hospitals include the external market competition, and the integration considers the cost of information sharing. Strategic priorities in a health care system also propose the integrated framework of competition and collaboration, with discussions on competition around information technology [28]. Furthermore, the influence of customer service–driven competition on horizontal integration through the mechanism of cost leadership is reflected in the service niche [15]. The outcome of vertical integration when facing customer service–driven competition comes through the one-stop shop service to achieve cost leadership [10]. Details of these three mechanisms are summarized in Table 1.

Table 1. Mechanisms from competition perceptions to integration plans.

Transaction cost economics theory		Integration plans	
Competition reasons	Mechanisms	Vertical integration *(economies of scope)*	Horizontal integration *(economies of scale)*
Uncertain environment	Hazards exchange	Expansion	Collaboration
Technology disruption	Transaction efficiency	Unified solutions	Delivery models
Customer service	Cost leadership	One-stop shop	Service niche

Methods

Data Collection

The effort to assess the linkage between competition and integration prospects of health systems is part of a broad project undertaken by the Health Administration Research Consortium at the Business School of the University of Colorado Denver [29]. The project involves an annual and broad study on health systems and collects insights via a survey of health system CEOs. The insights will help policymakers, practitioners, and academic stakeholders as they collaborate to create strategies to help the industry respond to the pandemic and prepare for the next crisis.

A survey questionnaire was developed in December 2020 to collect data from health systems and to study the environment that health systems face scientifically. The survey items came from prior literature, with questions reworded to fit the health systems context. Inputs were taken from researchers, consultants, and executives with appropriate expertise to design the questions. The survey was validated using a scientific process of expert evaluations and was pilot-tested with five top executives from the Health Administration Program Advisory Board. The survey questionnaire was revised and finalized in January 2021.

A contact list of CEOs was compiled from 624 health systems across the United States using data from multiple sources, contacts, professional networks, websites, and annual reports. The survey instrument was implemented in a professional survey platform and was mapped emails to the platform to create unique, trackable links for each health system. Email and phone solicitations were made in multiple rounds between January 25 and March 2, 2021. A total of 148 responses from the 624 CEOs contacted, representing a 23.7% response rate, out of which 13 incomplete responses could not be used, leaving 135 final usable responses.

The 135 health systems represented in this survey varied from 1 to 18 hospitals with 176 to 75,000 employees. The annual revenue in 2020 of the health systems ranged from US $0.7 million to US $14 billion. The health systems aggregately represented US $300 billion in revenues and 1.1 million employees across the United States.

We then matched the survey data set with secondary data collected from the Agency for Healthcare Research and Quality (AHRQ) compendium to glean a complete picture of the health systems. Finally, we had data from 124 health systems located across the United States. We analyzed this combined data set, which yielded several important insights.

Ethics Consideration

An ethics review was not applicable for this study. The data used was received through a leading professional consulting firm that anonymizes and provides secondary firm-level data for research and analysis to draw insights.

Variables and Measures

Table 2 describes the variables used in this study. The main variables in this study are vertical integration (VINT), horizontal integration (HINT), external environment uncertainty–related competition (EEUC), technology disruption–driven competition (TDDC), and customer service–driven competition (CSDC).

These variables were measured using 7-point Likert scales for relevant items. We tested the internal-consistency reliability of these multi-item variables using Cronbach α. The four α values were close to or greater than the recommended acceptable threshold of .70 for exploratory research.

Table 2. Description of variables, including survey questions and coding scheme.

Variable	Description
Integration plans	
VINT[a]	Develop through vertical integration
HINT[b]	Develop through horizontal integration
Competition perceptions	
Main question	What is needed for your health system to compete in today's post-COVID-19 economy?[c]
EEUC[d]	External environment uncertainty–relevant competition (Cronbach α=.84); focus on services in which you excel, reevaluate the business you are in, anticipate policy issues and be ready for that
TDDC[e]	Technology disruption–driven competition (Cronbach α=.73); transform through digital technologies, keeping current with technologies, new entrants disrupting the business model
CSDC[f]	Customer service–driven competition (Cronbach α=.67); loyalty of customers, develop a mix of talent, quality of services, and patient satisfaction
Contingent variables	
SIZE_B-SMALL, SIZE_B-MEDIUM, SIZE_B-LARGE	The three size variables of the health system are measured using the total beds managed by the health system across all hospitals, reported by the AHRQ[g] Hospital Compendium: SIZE_B_SMALL, <100 beds; SIZE_B_MEDIUM, 100-400 beds; SIZE_B_LARGE, >400 beds
REGION-NE, REGION-MW, REGION-SOUTH, REGION-WEST	The four region variables of the health systems are coded based on their primary location in the United States, following the Census Bureau categorization: REGION-NE, Northeast; REGION-MW, Midwest; REGION-SOUTH, South; REGION-WEST, West
TEACHING-NON, TEACHING-MINOR, TEACHING-MAJOR	The three teaching variables are coded based on the teaching status of a health system: TEACHING-NON, nonteaching; TEACHING-MINOR, minor teaching; TEACHING-MAJOR, major teaching
REVENUE-LOW, REVENUE-MEDIUM, REVENUE-HIGH	The three revenue variables of the health systems are measured using the health system's annual revenue across all hospitals: REVENUE-LOW, <US $2 billion; REVENUE-MEDIUM, US $2-5 billion; REVENUE-HIGH, >US $5 billion
HIGH-DSH-HOSP	The health system includes at least one high DSH[h] patient percentage hospital (1=yes, 0=no)
HIGH-BURDEN-SYS	Health system–wide uncompensated care burden flag (1=yes, 0=no)
HIGH-BURDEN-HOSP	The health system includes at least one high uncompensated care burden hospital (1=yes, 0=no)
OWNERSHIP	Predominantly investor-owned hospitals (1=yes, 0=no)
PHYSICIANS	The number of physicians in the health system, measured by the number of physicians reported by the AHRQ Hospital Compendium
HOSPITALS	Number of hospitals the health system has reported by the AHRQ Hospital Compendium

[a]VINT: vertical integration.

[b]HINT: horizontal integration.

[c]All questions are measured using a 7-point Likert scale; 1=strongly disagree to 7=strongly agree.

[d]EEUC: external environment uncertainty–related competition.

[e]TDDC: technology disruption–driven competition.

[f]CSDC: customer service–driven competition.

[g]AHRQ: Agency for Healthcare Research and Quality.

[h]DSH: discharge level.

The influencing factors examined in this study represent several categories: size, region, teaching status, revenue, and several other system characteristics. These variables are coded (see Table 2) to reflect the characteristics of a health system that may influence its competition perception and integration preference. Three size variables measure the number of beds across a health system (SIZE_B-SMALL, SIZE_B-MEDIUM, and SIZE_B-LARGE), four region variables reflect the location of a health system (REGION-NE, REGION-MW, REGION-SOUTH, and REGION-WEST), three teaching status–related variables assess the extent to which a health system is associated with a teaching program (TEACHING-NON, TEACHING-MINOR, and TEACHING-MAJOR). Three revenue variables measure the

annual revenue of a health system (REVENUE-LOW, REVENUE-MEDIUM, REVENUE-HIGH). In addition, we included variables to capture the discharge levels of patients (HIGH-DSH-HOSP), uncompensated care burden (HIGH-BURDEN-SYS and HIGH-BURDEN-HOSP), ownership status (OWNERSHIP), number of physicians (PHYSICIANS), and number of hospitals (HOSPITALS).

Sample Statistics

The descriptive statistics and pairwise correlations among the key variables used in this study appear in Table 3 and Multimedia Appendix 1, respectively. As shown in Table 3, health systems have a relatively high perception of TDDC compared with EEUC and CSDC. Furthermore, horizontal integration seems to be more popular for health systems than vertical integration.

In addition, to check for nonresponse bias, we compared the characteristics of responding and nonresponding health systems. Detailed comparisons are shown in Table 4. The *t* test results for all comparisons indicated no significant difference between respondents and nonrespondents.

Table 3. Summary statistics (N=124).

Variable[a]	Mean (SD)	Range
EEUC	5.10 (1.39)	1.67-7.00
TDDC	5.63 (0.94)	2.67-7.00
CSDC	5.12 (1.21)	2.00-6.67
VINT	4.51 (1.85)	1.00-7.00
HINT	4.97 (1.53)	1.00-7.00
SIZE_B-SMALL	0.09 (0.28)	0.00-1.00
SIZE_B-MEDIUM	0.37 (0.49)	0.00-1.00
SIZE_B-LARGE	0.54 (0.50)	0.00-1.00
REGION-NE	0.22 (0.42)	0.00-1.00
REGION-MW	0.24 (0.43)	0.00-1.00
REGION-SOUTH	0.35 (0.48)	0.00-1.00
REGION-WEST	0.18 (0.38)	0.00-1.00
TEACHING-NON	0.30 (0.46)	0.00-1.00
TEACHING-MINOR	0.48 (0.50)	0.00-1.00
TEACHING-MAJOR	0.22 (0.41)	0.00-1.00
REVENUE-LOW	0.61 (0.49)	0.00-1.00
REVENUE-MEDIUM	0.23 (0.43)	0.00-1.00
REVENUE-HIGH	0.15 (0.35)	0.00-1.00
HIGH-DSH-HOSP	0.33 (0.47)	0.00-1.00
HIGH-BURDEN-SYS	0.20 (0.40)	0.00-1.00
HIGH-BURDEN-HOSP	0.30 (0.46)	0.00-1.00
OWNSHIP	0.02 (0.13)	0.00-1.00
PHYSICIANS	1.84 (0.80)	1.00-3.00
HOSPITALS	1.50 (0.77)	1.00-3.00

[a]See Table 2 for a description of variable codes.

Table 4. Characteristics of responding and nonresponding health systems.

Characteristics[a]	Respondents (n=124), n (%)	Nonrespondents (n=511), n (%)	t test (633)
Size			
Small (6-99 beds)	11 (8.9)	42 (8.2)	–0.19
Medium (100-399 beds)	45 (36.3)	212 (41.3)	–0.56
Large (≥400 beds)	68 (54.8)	257 (50.3)	1.41
Region			
Northeast	27 (21.8)	117 (22.9)	0.07
Midwest	30 (24.2)	133 (26.0)	0.55
South	45 (36.3)	169 (33.1)	–0.48
West	22 (17.7)	92 (18.0)	–0.12
Physicians			
Small (51-199 physicians)	50 (40.3)	189 (37.0)	–0.74
Medium (200-999 physicians)	41 (33.1)	204 (39.9)	–0.69
Large (≥1000 physicians)	33 (26.6)	118 (23.1)	1.53
Hospitals			
Small (1-3 hospitals)	83 (66.9)	337 (65.9)	–1.27
Medium (4-6 hospitals)	20 (16.1)	67 (13.1)	–0.02
Large (≥7 hospitals)	21 (16.9)	107 (20.9)	0.81
Ownership status			
Investor-owned	3 (2.4)	15 (2.9)	–0.85
Noninvestor-owned	121 (97.6)	496 (97.1)	0.85
Teaching status			
Major teaching	29 (23.4)	138 (27.0)	–0.15
Minor teaching	58 (46.8)	225 (44.0)	–0.61
Nonteaching	37 (29.8)	148 (29.0)	0.85

[a]The number of physicians and hospitals are presented in this table in different categories for easy comparison across respondents and nonrespondents.

Statistical Analysis

To answer the two research questions stated earlier, we performed two sets of analyses. We used ordered logit regressions to estimate (1) the relationship of the three competition perceptions to specific hospital characteristics and (2) the relationship between competition perceptions and integration plans. The integration variables are ordinal (with a sequentially higher value), thus driving the decision for ordered logit regressions. The ordered logit approach does not assume equal intervals between levels in the dependent variable and is thus a preferred estimation than ordinary least square estimation processes that assumes equal linear intervals. The ordered logit model is as follows:

$$Y_i^* = \beta X_i + e_i,$$

where Y_i^* represents respondents' propensity to indicate higher levels of the dependent variables (ie, EEUC, TDDC, CSDC, VINT, and HINT), X_i is a set of explanatory variables, β is a vector of parameters, and e_i are disturbances.

We did not observe Y_i^*. Instead, we observed the ordinal dependent variable Y_i depending on the values of thresholds or cut-off points τ_{m-1} and τ_m. The probability distribution of Y_i is given as follows:

$$Pr(Y_i = m | X_i = F(\tau_m - X\beta) - F(\tau_{m-1} - X\beta)$$

Results

Estimation Outcomes

Table 5 shows the results of the ordered logit model estimation that describe the relationship between contingent factors and each of the three types of competition perceptions.

One teaching variable had a significant, negative association with EEUC. The major teaching health systems variable showed high statistical significance (*P*<.001). This suggests that major teaching health systems tend to have lower perceptions of EEUC. Based on the marginal-effects analysis, we found a 1.1% decrease in the probability of perceiving EEUC among major teaching health systems compared with nonteaching health systems.

There was also a significant negative relationship between high revenue and EEUC ($P<.001$), indicating that high-revenue health systems tend to perceive less EEUC than low-revenue health systems. The marginal-effects analysis suggested a 0.71% decrease in the probability of perceiving EEUC among high-revenue health systems than low-revenue health systems.

A high-burden system had a significant positive impact on EEUC ($P<.001$), while a high-burden hospital had a significant negative impact. These results indicate that (1) a health system with a system-wide high uncompensated care burden tends to perceive EEUC, while (2) a health system with no high uncompensated care burden hospital is less likely to perceive EEUC. We also examined the marginal effects of these two variables. The results indicated a 0.52% increase in the probability of perceiving EEUC by a health system with a system-wide high uncompensated care burden and a 1.32% decrease by a health system with at least one high uncompensated care burden hospital.

For TDDC, there was a significant negative relationship between this perception and both medium size ($P<.001$) and large size ($P=.04$), indicating that smaller-sized health systems are more likely to have a higher level of TDDC. The marginal-effects analysis showed that the probability changes for these two factors are 3.08% and 2.13%, respectively.

Region had significant effects on the perception of TDDC. For example, health systems in the south ($P<.001$) have a higher TDDC perception than those in the northeast. The change in the marginal effects was 0.92%. Similarly, there was a significant positive relationship between high revenue and TDDC ($P<.001$). This result indicates that high-revenue health systems tend to perceive more TDDC than low-revenue health systems. The marginal effect of this change was 0.93%.

The relationship between a high system-wide burden and TDDC and the relationship between total hospitals in a health system and TDDC were both significant and negative ($P<.001$). These results suggest that health systems without a system-wide high uncompensated care burden and those with fewer hospitals are more likely to perceive TDDC. The marginal effects for these two variables were 1.14% and 0.59%, respectively.

For CSDC, compared with small-sized health systems (ie, those with a fewer number of beds), medium- and large-sized health systems ($P<.001$) are less likely to perceive CSDC. These results reveal the strong influence of health systems' size on their competition perceptions regarding customer service provision. More specifically, marginal-effects analysis indicated a 1.11% and 0.49% decrease in the probability of perceiving CSDC for medium- and large-sized health systems, respectively.

Table 5. Influence of contingent factors from the ordered logit model estimation.[a]

Variables	EEUC[b]		TDDC[c]		CSDC[d]	
	Coefficient (SE)	P value	Coefficient (SE)	P value	Coefficient (SE)	P value
SIZE-MEDIUM	−.379 (.122)	.002	−1.789 (.184)	<.001	−1.468 (.124)	<.001
SIZE-LARGE	.463 (.358)	.20	−1.673 (.826)	.04	−.864 (.164)	<.001
REGION-MW	.046 (.261)	.86	.011 (.277)	.97	−.660 (.484)	.17
REGION-SOUTH	.292 (.099)	.003	.824 (.118)	<.001	.294 (.378)	.44
REGION-WEST	1.210 (.717)	.09	.210 (.570)	.71	.773 (.959)	.42
TEACHING-MINOR	−.575 (.548)	.29	.304 (.183)	.10	−.181 (.424)	.67
TEACHING-MAJOR	-1.348 (.068)	<.001	−.081 (.211)	.70	−.995 (.112)	<.001
REVENUE-MEDIUM	−.702 (.449)	.12	.932 (.359)	.009	−.217 (.054)	<.001
REVENUE-HIGH	−.961 (.162)	<.001	1.056 (.130)	<.001	−.200 (.106)	.06
HIGH-DSH-HOSP	.431 (.848)	.61	.639 (.282)	.02	.425 (.422)	.31
HIGH-BURDEN-SYS	1.434 (.056)	<.001	−.739 (.116)	<.001	.314 (.649)	.63
HIGH-BURDEN-HOSP	−1.668 (.129)	<.001	.241 (.169)	.15	−.977 (.311)	.002
OWNERSHIP	.333 (.552)	.55	−1.644 (1.337)	.22	−.667 (1.477)	.65
PHYSICIANS	.224 (.107)	.04	−.022 (.643)	.97	−.057 (.205)	.78
HOSPITALS	.004 (.028)	.88	−.482 (.125)	<.001	.221 (.130)	.09
Pseudo R^2	0.057	—[e]	0.036	—	.043	—

[a]The results of the cut-off points are omitted for brevity.

[b]EEUC: environment uncertainty–related competition.

[c]TDDC: technology disruption–driven competition.

[d]CSDC: customer service–driven competition.

[e]not applicable.

XSL•FO

RenderX

There was a significant negative relationship between major teaching health systems and CSDC ($P<.001$), indicating that compared with nonteaching health systems, those with a teaching focus tend to perceive less CSDC. The marginal-effects analysis indicated a 0.78% decrease in the probability of perceiving CSDC among major teaching health systems than among nonteaching health systems.

In addition, there was a significant negative relationship between CSDC and medium revenue ($P<.001$), suggesting that health systems with a midrange revenue may not favor perceived CSDC. The marginal-effects analysis showed a 0.13% impact for this revenue variable.

Table 6 and Table 7 display the estimation results for the second set of models, illustrating the relationship between competition perceptions and vertical and horizontal integration, respectively.

The results in Table 6 show different relationships between the three types of competition perceptions and vertical integration. Models 1-3 demonstrate the direct relationships of EEUC, TDDC, and CSDC with VINT, respectively, while model 4 is the combined model, which includes all three competition perceptions. The results indicated a significant positive relationship ($P<.001$) between EEUC and VINT, and a significant negative relationship ($P<.001$) between TDDC and VINT. The relationship between CSDC and VINT was not significant. More specifically, the results suggest that health systems with higher EEUC perceptions tend to choose vertical integration plans. The marginal-effects analysis showed a value of 4.64% for this variable. However, health systems with a higher level of TDDC are less likely to opt for vertical integration. The probability change, according to the marginal-effects analysis, was 4.11% for TDDC.

Table 6. Ordered logit model estimation results for competition types and vertical integration.[a]

Variables	Model 1		Model 2		Model 3		Model 4 (combined)	
	Coefficient (SE)	P value	Coefficient (SE)	P value	Coefficient (SE)	P value	Coefficient (SE)	P value
EEUC[b]	1.203 (.152)	<.001	—[c]	—	—	—	1.601 (.255)	<.001
TDDC[d]	—	—	−.515 (.132)	<.001	—	—	−1.029 (.162)	<.001
CSDC[e]	—	—	—	—	.163 (.183)	.37	-.111 (.161)	.49
SIZE	.228 (.666)	.73	.288 (.284)	.31	.371 (.518)	.47	.021 (.371)	.95
REGION	.451 (.205)	.03	.553 (.015)	<.001	.460 (.030)	<.001	.567 (.141)	<.001
OWNERSHIP	−.135 (3.608)	.97	.347 (2.488)	.89	.561 (3.942)	.89	−.612 (1.908)	.75
TEACHING	.648 (.587)	.27	−.053 (.493)	.92	.026 (.602)	.97	.739 (.606)	.22
REVENUE	−.105 (.042)	.01	−.288 (.035)	<.001	−.349 (.033)	<.001	.111 (.066)	.09
HIGH-DSH[f]-HOSP[g]	−.793 (.635)	.21	−.184 (.447)	.68	−.453 (.301)	.13	−.590 (.607)	.33
HIGH-BURDEN-SYS[h]	.149 (.958)	.88	1.002 (.434)	.02	.924 (.636)	.15	−.026 (.851)	.98
HIGH-BURDEN-HOSP	−.529 (.408)	.20	−1.365 (.178)	<.001	−1.087 (.354)	.002	−.560 (.303)	.07
PHYSICIANS	−.347 (.211)	.10	.010 (.176)	.95	−.035 (.160)	.83	−.282 (.437)	.52
HOSPITALS	.022 (.221)	.92	−.048 (.100)	.63	−.020 (.081)	.80	−.086 (.294)	.77
Pseudo R^2	0.187	N/A[i]	0.069	N/A	.053	N/A	.254	N/A
Mean VIF[j]	1.89	N/A	1.73	N/A	1.73	N/A	2.07	N/A

[a]The results of the cut-off points are omitted for brevity.

[b]EEUC: environment uncertainty–related competition.

[c]Not included in model.

[d]TDDC: technology disruption–driven competition

[e]CSDC: customer service–driven competition.

[f]DSH: discharge.

[g]HOSP: hospital.

[h]SYS: system.

[i]N/A: not applicable.

[j]VIF: variance inflation factor.

Similarly, the results in Table 7 display the different direct (models 1, 2, and 3) and combined (model 4) effects of the three types of competition perceptions on horizontal integration. The most significant relationship was the impact of EEUC on HINT ($P<.001$). This significant and positive relationship is consistent in both the direct and combined models, indicating that health

systems with higher EEUC perceptions have a higher probability of following horizontal integration plans. The negative relationship between TDDC and HINT was significant (*P*=.05) in the combined model. This result suggests a lower probability of adopting horizontal integration for health systems with higher TDDC perceptions. Finally, although the direct relationship between CSDC and HINT was not significant in the direct model, the combined model showed a highly significant negative relationship (*P*<.001) between these variables. This result indicates that health systems that perceive a higher level of CSDC are more likely to pursue a horizontal integration approach.

Table 7. Ordered logit model estimation results for competition types and horizontal integration.[a]

Variables	Model 1		Model 2		Model 3		Model 4 (combined)	
	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value
EEUC[b]	.619 (.036)	<.001	—[c]	—	—	—	1.360 (.130)	<.001
TDDC[d]	—	—	–.655 (.336)	.05	—	—	–.219 (.111)	.05
CSDC[e]	—	—	—	—	–.309 (.289)	.28	–1.153 (.210)	<.001
SIZE	.395 (.422)	.35	.305 (.258)	.24	.423 (.443)	.34	.365 (.135)	.007
REGION	.199 (.230)	.39	.359 (.097)	<.001	.372 (.041)	<.001	.397 (.075)	<.001
OWNERSHIP	–.712 (1.913)	.71	–.666 (1.375)	.63	–.539 (1.543)	.73	–1.679 (.829)	.04
TEACHING	.152 (.696)	.83	–.041 (.774)	.96	–.135 (.817)	.87	.035 (.669)	.96
REVENUE	–.138 (.274)	.61	–.226 (.263)	.39	–.432 (.120)	<.001	–.120 (.373)	.75
HIGH-DSH[f]-HOSP[g]	–.653 (.317)	.04	–.334 (.227)	.14	–.375 (.255)	.14	–.474 (.223)	.03
HIGH-BURDEN-SYS[h]	.947 (.395)	.02	1.241 (.490)	.01	1.440 (.470)	.002	.740 (.648)	.25
HIGH-BURDEN-HOSP	–.199 (.513)	.70	–.710 (.454)	.12	–1.019 (.228)	<.001	–.548 (.378)	.15
PHYSICIANS	.177 (.165)	.28	.224 (.313)	.47	.310 (.184)	.09	.183 (.300)	.54
HOSPITALS	–.219 (.507)	.67	–.193 (.468)	.68	–.048 (.321)	.88	–.004 (.498)	.99
Pseudo R^2	.102	N/A[i]	.080	N/A	.062	N/A	.212	N/A
Mean VIF[j]	1.78	N/A	1.73	N/A	1.74	N/A	2.00	N/A

[a]The results of the cut points are omitted for brevity.

[b]EEUC: environment uncertainty–related competition.

[c]Not included in model.

[d]TDDC: technology disruption–driven competition

[e]CSDC: customer service–driven competition.

[f]DSH: dispatch.

[g]HOSP: hospital.

[h]SYS: system.

[i]N/A: not applicable.

[j]VIF: variance inflation factor.

Discussion

Principal Findings

This study explored the differences of three competition perceptions of health systems in the United States, contingent on their characteristics. The study further examined how competition aspects are related to the different integration plans. The main findings are two-fold. For the first set of explorations on the main factors that influence health systems' competition perceptions, asking about the similarities and differences of such influences and their implications, we found that size, teaching status, revenue, and burden are the four main factors influencing health systems' competition perceptions.

First, the results indicate that small-sized health systems will perceive a higher level of all three types of competition perceptions. Therefore, a driving reason for integration plans among small-sized health systems may be to mitigate competition. The influences of size on all three competition perceptions were consistent.

Second, compared with nonteaching health systems, major teaching health systems perceive less EEUC and CSDC. A potential reason is an emphasis on major teaching health

systems; they are less bothered by the external environment and customer service provision changes.

Third, the results indicate that revenue has opposite influences for TDDC perceptions and EEUC and CSDC perceptions. On the one hand, the results suggest that high-revenue health systems perceive a higher level of TDDC. On the other hand, health systems with more revenue (both medium and high revenue) perceive lower EEUC and CSDC. These results demonstrate the nuanced influences of revenue on competition perceptions, and imply that the role of revenue should be examined more carefully in relevant decision-making.

Fourth, according to prior research (AHRQ), there are two types of uncompensated care burdens: a system-wide burden and hospital-level burden. This study employed two variables, high-burden system and high-burden-hospital, to capture these two types of burdens. The results showed that they have different influences on different competition perceptions. For example, health systems with a system-wide burden are more likely to perceive EEUC, while those with at least one high-burden hospital are less likely to perceive EEUC.

For the second set of examinations, we found a consistent significant positive relationship between EEUC and both vertical and horizontal integration; a consistent negative relationship, with changing significance, between TDDC and the two integration plans; and a negative relationship between CSDC and horizontal integration (see Table 8).

The results suggest the following. First, the strongest motivation for integration or an M&A is EEUC perceptions. When there is external environment uncertainty, organizations tend to pool their efforts to overcome the difficulties. Second, counterintuitively, TDDC perceptions hinder integrations. Data, privacy, and intellectual property issues may be the underlying reasons for such findings. For example, Internet of Things (IoT) technology can bring in a privacy threat in the process of integration. There will be data acquisition, aggregation, and integration in an IoT ecosystem where different systems build the linkage [30]. Third, to provide better customer services, health systems may not pursue a vertical integration strategy (it is unclear in this study), but they will not pursue horizontal integration. Organizations may not think they can gain better customer services by acquiring a similar company.

Table 8. Summary of findings: relationship between competition and integration.

Variable	Vertical integration	Horizontal integration
EEUC[a]	Positive***	Positive***
TDDC[b]	Negative***	Negative*
CSDC[c]	N/A[d]	Negative***

[a]EEUC: environment uncertainty–related competition.

[b]TDDC: technology disruption–driven competition.

[c]CSDC: customer service–driven competition.

*** $P<.001$,* $P<.05$

Implications

The findings of this study have several practical and policy implications. First, this study sheds light on the influence of the characteristics of health systems on competition perceptions. Medium size, major teaching status, high revenue, and having at least one uncompensated care burden hospital lead to lower perceptions of EEUC. Health systems with a high system-wide burden of uncompensated care feel a stronger sense of competition due to external environment uncertainty. Such differences support our definition and understanding of EEUC perceptions. In other words, external environment uncertainty is considered at the health-system level. Furthermore, the system-wide burden has an opposite influence on EEUC and TDDC. Health systems with a system-wide burden tend to perceive a higher level of external environment–driven competition but a lower level of technology-driven competition. This is a meaningful finding. The implication is that a system-wide burden motivates or even forces health systems to worry about external environment uncertainty, including changes in Medicare policy. Health systems with a system-wide uncompensated care burden may not think technological disruption can solve their problems. The main takeaway is that even in the post-COVID-19 era, where technologies such as

telehealth dominate and change the health care industry, other organizational factors are still critical for competition.

Recent research provides a rationale in the technology context. For example, digital orientations of smaller and less complex health systems are not aspirational [31]. This study also implies that medium size, system-wide burden, and more hospitals in a system can have negative influences on TDDC, while health systems in the south and those with high revenue perceive TDDC negatively at an even higher level. In other words, health systems with such characteristics should leverage the technologies for their organizations. Furthermore, revenue from customer services may not be equally important for all health systems when examining the influence of CSDC. This supports the prior literature on competitive markets [32,33]. Thus, policies designed to reform the delivery system must consider the diversity in system characteristics [17].

Second, the critical implication of this study is related to how EEUC, TDDC, and CSDC perceptions influence vertical and horizontal integration plans by the leaders of health systems. We found that the EEUC perception is influential on both vertical and horizontal integration plans. This finding indicates that in unstable markets and situations such as the COVID-19–relevant financial disruptions, there may be more

M&As. As a result, policymakers and practitioners may want to get ready for the changes in the post-COVID-19 new normal.

Regarding the influence of TDCC perceptions on vertical and horizontal integration, we found that technology-based competition drives less vertical integration. Given that health systems have recently aspired to accommodate remote- and virtual-care delivery options using technology, this is surprising. Plausibly, rather than acquiring firms, they may be focusing on developing in-house expertise. Indeed, a few telehealth M&As have been at the technology provider level rather than at the health system level. This may signal that health systems are preparing to develop technological expertise rather than making it an excuse for quick buyouts or mergers.

In addition, customer services should be an essential parameter for all health systems. However, whether it drives integration decisions remains an open question. We found that CSDC perceptions are not a strong driver of integration, although the external environment and technological uncertainties do affect horizontal integration. In other words, in a highly competitive market that includes external-, technological-, and customer service–relevant competitions, health systems will pursue horizontal integration strategies through mergers. This reflects an opportunistic behavior to monopolize markets and lock customers into one system.

There are a few other findings relevant to the contingent factors that influence integrations. For example, we found that health systems in specific regions and those with high revenue and high burden have a higher propensity for vertical integration, with slight variations in these findings for horizontal integration. We do not tease out many implications from these findings, except that they suggest that M&A decisions are influenced by a set of factors other than simply competition aspects.

There are strong indications that external uncertainty is a significant cause for thinking about integration. Perhaps policy interventions for fiscal assurance are a way to help suffering health systems survive the pandemic. The aspirations of some healthy and large health systems to take advantage of this situation to buy out weaker ones and emerge as the key dominant player in a region need to be kept in check.

Limitations and Directions for Further Research

This study has a few limitations that future studies may be able to address. Focusing on internal issues (eg, management and coordination, value-chain dynamics) and demand-driven integration plans may be other perspectives to explore for health systems. Furthermore, opportunities for and barriers to integration may be studied in greater detail. Recent studies have shown that hospital M&As are associated with decreased patient experience measures, along with no evidence of quality improvement with the change in ownership of health entities [20]. Studies on the association between structural changes have yielded mixed results on cost, quality, and patient experience outcomes, indicating that the structural integration of health care organizations is conceptually distinct from integrated care delivery. Structural integration may or may not lead to integrated care, and providers and policymakers should focus on the conditions and strategies that enable structurally integrated

organizations to capitalize on their ability to deliver more integrated care [16].

There are also concerns that integrated systems may increase prices and cost of care without commensurate improvements in quality and outcomes. The vertical integration approach showed no differences or lower efficiencies, as measured by utilization, spending, and prices [34]. Some evidence indicates that higher levels of integration are associated with higher levels of health care spending and increasing prices [35]. Other evidence shows mixed results, suggesting no significant effects or effects dependent on insurance type. While increased efficiencies may be possible, emerging research raises concerns about anticompetitive behavior, spending increases, and uncertain effects on quality. Vertical integration poses a threat to the affordability of health services and merits special attention from policymakers and antitrust authorities [13].

Reviews of hospital markets have found that concentrated markets are associated with higher hospital prices, with price increases often exceeding 20% after mergers. Of even more significant concern, reviews find that these price increases do not improve quality. In some cases, higher hospital concentration is associated with higher mortality rates. The higher concentration is associated with higher physician prices across a range of services. Despite evidence of associated price increases with no significant quality or efficiency improvements, both vertical and horizontal integration approaches have proceeded unchallenged. Health care consolidation concerns policymakers and regulators because market concentration can harm patients by increasing prices and premiums without accompanying improved care quality.

We also recognize that integration is not insufficient by itself, although we acknowledge that our underlying tone in this study has been to avoid overly integrated monopolistic endeavors. Future work might investigate M&As with outcomes, as done in prior work [9,26]. It is also worth noting that increases in the market concentration of health care providers and insurers have been examined nationally, and that increases in market concentration are associated with increases in prices and premiums. However, local markets for health care can differ dramatically. At the state level, laws and regulations, and the mix of providers and insurers, make markets in each state vastly different [36].

Another limitation of this study is that we used the cross-sectional data set to examine the relationships between variables. Future studies are planned to collect and use multiple years of data from health systems to address this limitation and provide causal inferences.

Conclusions

Disruptions in health care are an emerging challenge and stem from a variety of sources. Recently, the COVID-19 pandemic has been a significant factor. In addition, technological disruptions and demands from empowered customers have put enormous pressure on health systems to shape different strategies for survival. Integration is one such strategy. However, understanding why leaders plan for integration is an essential insight for the health care sector. This study unravels

competition-integration dynamics, and relates external environment uncertainty, technological competition, and customer services–driven competition to vertical and horizontal integration plans. Almost all health systems have some plans for integrations [37]; however, we found that environmental uncertainty drives integration more than other competitive factors. In addition, health systems with heavy competition dynamics will opt for mergers to alleviate survival challenges.

Competition-driven integration is unavoidable. However, overly integrated markets may lead to monopolistic entities and behavior, and that potential needs to be carefully managed and avoided with policy-level interventions. While interventions in the US health care sector are achieved through laws and regulations, proactively managing competition is an essential aspect of shaping policy interventions, and requires broader discussion and action.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1
Pairwise correlations among main variables (N=124).
[DOCX File , 20 KB - formative_v6i3e32477_app1.docx]

References

1. Kaufman H. Hospital merger and acquisition activity continues to rise, according to Kaufman Hall analysis. Cision. 2018 Jul 12. URL: https://www.prnewswire.com/news-releases/hospital-merger-and-acquisition-activity-continues-to-rise-according-to-kaufman-hall-analysis-300679914.html [accessed 2021-10-23]
2. LaPointe J. Healthcare merger, acquisition deal value increased 145% in 2017. RevCycleIntelligence. 2018 Jan 29. URL: https://revcycleintelligence.com/news/healthcare-merger-acquisition-deal-value-increased-145-in-2017 [accessed 2021-10-23]
3. Han PKJ, Babrow A, Hillen MA, Gulbrandsen P, Smets EM, Ofstad EH. Uncertainty in health care: Towards a more systematic program of research. Patient Educ Couns 2019 Oct;102(10):1756-1766. [doi: 10.1016/j.pec.2019.06.012] [Medline: 31227333]
4. Henry Ford Health System, Michigan State partner to improve access, outcomes and address health inequities, disparities. Henry Ford Health System. 2021 Jan 25. URL: https://www.henryford.com/news/2021/01/henry-ford-and-msu-finalize-partnership [accessed 2021-10-23]
5. LaPointe J. Atrium Health, Wake Forest Baptist Health finalize merger. RevCycleIntelligence. 2020 Oct 12. URL: https://revcycleintelligence.com/news/atrium-health-wake-forest-baptist-health-finalize-merger [accessed 2021-10-23]
6. Pennic F. SPAC merges with 2 telehealth companies to form public digital health company in $1.35B deal. HIT Consultant. 2020 Nov 23. URL: https://hitconsultant.net/2020/11/23/spac-telehealth-merge-form-digital-health-company/#.YNNYs2hKgvg [accessed 2021-10-23]
7. Bryant B. Cimarron Healthcare Capital completes acquisition of Ascent Behavioral Health. Behavioral Health Business. 2020 Feb 19. URL: https://bhbusiness.com/2020/02/19/cimarron-healthcare-capital-completes-acquisition-of-ascent-behavioral-health/ [accessed 2021-10-23]
8. Kindred Healthcare and Landmark Medical Center announce definitive agreement forming joint venture to own and operate Rehabilitation Hospital of Rhode Island. Business Wire. URL: https://www.businesswire.com/news/home/20200617005402/en/Kindred-Healthcare-Landmark-Medical-Center-Announce-Definitive [accessed 2021-10-23]
9. Beaulieu ND, Dafny LS, Landon BE, Dalton JB, Kuye I, McWilliams JM. Changes in quality of care after hospital mergers and acquisitions. N Engl J Med 2020 Jan 02;382(1):51-59 [FREE Full text] [doi: 10.1056/NEJMsa1901383] [Medline: 31893515]
10. Gallacher K, May CR, Montori VM, Mair FS. Understanding patients' experiences of treatment burden in chronic heart failure using normalization process theory. Ann Fam Med 2011;9(3):235-243 [FREE Full text] [doi: 10.1370/afm.1249] [Medline: 21555751]

XSL•FO
RenderX

11. Fulop N, Protopsaltis G, King A, Allen P, Hutchings A, Normand C. Changing organisations: a study of the context and processes of mergers of health care providers in England. Soc Sci Med 2005 Jan;60(1):119-130. [doi: 10.1016/j.socscimed.2004.04.017] [Medline: 15482872]

12. Thaldorf C, Liberman A. Integration of health care organizations: using the power strategies of horizontal and vertical integration in public and private health systems. Health Care Manag 2007;26(2):116-127. [doi: 10.1097/01.HCM.0000268614.41115.fc] [Medline: 17464224]

13. Post B, Buchmueller T, Ryan AM. Vertical integration of hospitals and physicians: economic theory and empirical evidence on spending and quality. Med Care Res Rev 2018 Aug;75(4):399-433. [doi: 10.1177/1077558717727834] [Medline: 29148355]

14. Axelsson R, Axelsson SB. Integration and collaboration in public health--a conceptual framework. Int J Health Plann Manage 2006;21(1):75-88. [doi: 10.1002/hpm.826] [Medline: 16604850]

15. Staňková P, Papadaki Š, Dvorský J. The reducing hospital costs through horizontal integration. Iran J Public Health 2019 Nov;48(11):2016-2024 [FREE Full text] [Medline: 31970100]

16. Kerrissey MJ, Clark JR, Friedberg MW, Jiang W, Fryer AK, Frean M, et al. Medical group structural integration may not ensure that care is integrated, from the patient's perspective. Health Aff (Millwood) 2017 May 01;36(5):885-892. [doi: 10.1377/hlthaff.2016.0909] [Medline: 28461356]

17. Furukawa MF, Machta RM, Barrett KA, Jones DJ, Shortell SM, Scanlon DP, et al. Landscape of health systems in the United States. Med Care Res Rev 2020 Aug;77(4):357-366 [FREE Full text] [doi: 10.1177/1077558718823130] [Medline: 30674227]

18. Armitage GD, Suter E, Oelke ND, Adair CE. Health systems integration: state of the evidence. Int J Integr Care 2009 Jun 17;9(2):e82 [FREE Full text] [doi: 10.5334/ijic.316] [Medline: 19590762]

19. O'Hanlon CE. Impacts of health care industry consolidation in Pittsburgh, Pennsylvania: a qualitative study. Inquiry 2020;57:46958020976246 [FREE Full text] [doi: 10.1177/0046958020976246] [Medline: 33233979]

20. Li J, Zhang Y, Ma L, Liu X. The impact of the internet on health consultation market concentration: an econometric analysis of secondary data. J Med Internet Res 2016 Oct 28;18(10):e276 [FREE Full text] [doi: 10.2196/jmir.6423] [Medline: 27793793]

21. Fisher ES, Shortell SM, O'Malley AJ, Fraze TK, Wood A, Palm M, et al. Financial integration's impact on care delivery and payment reforms: a survey of hospitals and physician practices. Health Aff (Millwood) 2020 Aug;39(8):1302-1311 [FREE Full text] [doi: 10.1377/hlthaff.2019.01813] [Medline: 32744948]

22. Donato R. Extending transaction cost economics: towards a synthesised approach for analysing contracting in health care markets with experience from the Australian private sector. Soc Sci Med 2010 Dec;71(11):1989-1996. [doi: 10.1016/j.socscimed.2010.09.016] [Medline: 20943302]

23. Schniederjans DG, Hales DN. Cloud computing and its impact on economic and environmental performance: a transaction cost economics perspective. Decision Support Systems 2016 Jun;86:73-82. [doi: 10.1016/j.dss.2016.03.009]

24. Judge WQ, Dooley R. Strategic alliance outcomes: a transaction-cost economics perspective. Br J Management 2006 Mar;17(1):23-37. [doi: 10.1111/j.1467-8551.2005.00441.x]

25. Porter ME. Competitive advantage: creating and sustaining superior performance. New York, NY: Free Press; 1985.

26. Blair R, Durrance C, Sokol D. Hospital mergers and economic efficiency. Wash L Rev 2016;91(1):1-70.

27. Alexander JA, Burns LR, Morrisey MA, Johnson V. CEO perceptions of competition--and strategic response in hospital markets. Med Care Res Rev 2001 Jun 19;58(2):162-93; discussion 229. [doi: 10.1177/107755870105800202] [Medline: 11398645]

28. Brown AD, Alikhan LM, Sandoval GA, Seeman N, Baker GR, Pink GH. Acute care hospital strategic priorities: perceptions of challenges, control, competition and collaboration in Ontario's evolving healthcare system. Healthc Q 2005;8(3):36-47. [Medline: 16078398]

29. Health Systems Climate Study Report. CU Denver Business School. URL: https://business.ucdenver.edu/harc/2021ClimateStudyReport [accessed 2021-10-23]

30. Madaan N, Ahad MA, Sastry SM. Data integration in IoT ecosystem: information linkage as a privacy threat. Comput Law Secur Rev 2018 Feb;34(1):125-133. [doi: 10.1016/j.clsr.2017.06.007]

31. Khuntia J, Ning X, Stacey R. Digital orientation of health systems in the post-COVID-19 "new normal" in the United States: cross-sectional survey. J Med Internet Res 2021 Aug 16;23(8):e30453 [FREE Full text] [doi: 10.2196/30453] [Medline: 34254947]

32. Vallabhajosyula S, Patlolla SH, Dunlay SM, Prasad A, Bell MR, Jaffe AS, et al. Regional variation in the management and outcomes of acute myocardial infarction with cardiogenic shock in the United States. Circ Heart Fail 2020 Feb;13(2):e006661 [FREE Full text] [doi: 10.1161/CIRCHEARTFAILURE.119.006661] [Medline: 32059628]

33. DesRoches CM, Worzala C, Joshi MS, Kralovec PD, Jha AK. Small, nonteaching, and rural hospitals continue to be slow in adopting electronic health record systems. Health Aff (Millwood) 2012 May;31(5):1092-1099. [doi: 10.1377/hlthaff.2012.0153] [Medline: 22535503]

34.   Machta RM, Maurer KA, Jones DJ, Furukawa MF, Rich EC. A systematic review of vertical integration and quality of care, efficiency, and patient-centered outcomes. Health Care Manage Rev 2019;44(2):159-173. [doi: 10.1097/HMR.0000000000000197] [Medline: 29613860]

35.   Kelsall A, Schuman A, Martin M, Dimick M. The impact of health system mergers on quality and cost of care (Master Class: Health Systems). Michigan J Med 2019;4(2):1. [doi: 10.3998/mjm.13761231.0004.209]

36.   Scheffler RM, Arnold DR, Whaley CM. Consolidation trends In California's health care system: impacts on ACA premiums and outpatient visit prices. Health Aff (Millwood) 2018 Sep;37(9):1409-1416. [doi: 10.1377/hlthaff.2018.0472] [Medline: 30179552]

37.   Fulton BD. Health care market concentration trends in the United States: evidence and policy responses. Health Aff (Millwood) 2017 Sep 01;36(9):1530-1538. [doi: 10.1377/hlthaff.2017.0556] [Medline: 28874478]

Abbreviations

AHRQ: Agency for Healthcare Research and Quality
CEO: chief executive officer
CSDC: customer service driven competition
EEUC: external environment uncertainty relevant competition
HARC: Health Administration Research Consortium
IoT: Internet of Things
HINT: horizontal integration
M&A: merger and acquisition
TDDC: technology disruption driven competition
VINT: vertical integration

Original Paper

The Successes and Challenges of Implementing Telehealth for Diverse Patient Populations Requiring Prenatal Care During COVID-19: Qualitative Study

Ruth Farrell[1,2,3], MD, MA; Christina Collart[1], MEd; Caitlin Craighead[1], MPH; Madelyn Pierce[1], BS; Edward Chien[1], MD, MBA; Richard Frankel[4], PhD; Brownsyne Tucker Edmonds[5], MD, MPH; Uma Perni[1], MD; Marissa Coleridge[1,2], MS; Angela C Ranzini[6], MD; Susannah Rose[7], PhD

[1]Obstetrics and Gynecology and Women's Health Institute, Cleveland Clinic, Cleveland, OH, United States

[2]Genomic Medicine Institute, Cleveland Clinic, Cleveland, OH, United States

[3]Center for Bioethics, Cleveland Clinic, Cleveland, OH, United States

[4]Research Innovation and Education, Cleveland Clinic Lerner College of Medicine, Cleveland, OH, United States

[5]Department of Obstetrics and Gynecology, Indiana University, Indianapolis, IN, United States

[6]Department of Obstetrics and Gynecology, MetroHealth Medical Center, Cleveland, OH, United States

[7]Center for Patience Experience, Cleveland Clinic, Cleveland, OH, United States

Corresponding Author:
Ruth Farrell, MD, MA
Obstetrics and Gynecology and Women's Health Institute
Cleveland Clinic
9500 Euclid Ave
Cleveland, OH, 44195
United States
Phone: 1 216 445 7085
Email: farrelr@ccf.org

Abstract

Background: Although telehealth appears to have been accepted among some obstetric populations before the COVID-19 pandemic, patients' receptivity and experience with the rapid conversion of this mode of health care delivery are unknown.

Objective: In this study, we examine patients' prenatal care needs, preferences, and experiences during the COVID-19 pandemic, with the aim of supporting the development of successful models to serve the needs of pregnant patients, obstetric providers, and health care systems during this time.

Methods: This study involved qualitative methods to explore pregnant patients' experiences with prenatal health care delivery at the onset of the COVID-19 pandemic. We conducted in-depth interviews with pregnant patients in the first and second trimester of pregnancy who received prenatal care in Cleveland, Ohio, from May to July 2020. An interview guide was used to probe experiences with health care delivery as it rapidly evolved at the onset of the pandemic.

Results: Although advantages of telehealth were noted, there were several concerns noted with the broad implementation of telehealth for prenatal care during the pandemic. This included concerns about monitoring the pregnancy at home; the need for additional reassurance for the pregnancy, given the uncertainties presented by the pandemic; and the ability to have effective patient-provider discussions via a telehealth visit. The need to tailor telehealth to prenatal health care delivery was noted.

Conclusions: Although previous studies have demonstrated that telehealth is a flexible and convenient alternative for some prenatal appointments, our study suggests that there may be specific needs and concerns among the diverse patient groups using this modality during the pandemic. More research is needed to understand patients' experiences with telehealth during the pandemic and develop approaches that are responsive to the needs and preferences of patients.

XSL•FO
RenderX

Introduction

Telehealth is a rapidly evolving approach to the delivery of health care, and significant strides have been made in its use for the delivery of prenatal care [1-4]. The appeal of telehealth is growing, as it provides not only a convenient and cost-effective way to deliver prenatal care, but also a way to improve patient experience [1,5-7]. Although telehealth was incrementally implemented before the emergence of COVID-19 across different patient populations and specialties of medicine, the pandemic radically altered both the timeline and the trajectory of its use to prevent the spread of the virus among health care providers, patients, and communities.

Preventing the spread of COVID-19 has been a particular priority in delivering prenatal care, because prior infectious disease threats, such as H1N1 and Zika, have had major implications for pregnant patients and infants [8-10]. In addition, as there is limited information about the impact of SARS-CoV-2 on pregnant women and newborns, health care institutions made tremendous efforts to ensure social distancing and infection prevention among this population [11-15]. Consequently, there has been rapid adoption of telehealth across diverse obstetric patient populations and health care systems [3].

As a result of initial efforts to implement telehealth in prenatal care, it was possible to begin to operationalize telehealth strategies that allowed patients to continue accessing health care during the pandemic, while eliminating the exposure risks of presenting to an outpatient center or hospital setting. Yet, this required instituting widespread acceptance of this approach among obstetric patient and provider populations who otherwise might not have embraced this type of visit before. For example, some providers either may not have used or elected to use telehealth prior to the pandemic. In addition, some patients may have had limited access to the equipment, internet services, or other resources needed to use telehealth; such issues have often been observed among patients from minority or low-socioeconomic groups [16,17]. Moreover, some patients may not have had an insurance plan that would cover the cost of a telehealth visit, while others may have elected not to use telehealth services because of a preference for in-person visits.

Consequently, although telehealth appears to have been accepted among some obstetric populations before the pandemic, patients' receptivity and experience with the rapid conversion to telehealth are unknown. Furthermore, it is unknown how patients will respond to telehealth visits for obstetric management amidst all the uncertainties presented by the pandemic. It is critical to understand not only patients' experiences with this mode of prenatal health care delivery during the pandemic but also how those experiences may influence health care delivery in the future. Therefore, we conducted a study to understand patients' prenatal care needs, preferences, and experiences during the COVID-19 pandemic to support the development of successful models to serve the needs of pregnant patients, obstetric providers, and health care systems during this time.

Methods

This study was developed as a qualitative study to explore emerging concepts and themes as they relate to obstetric health care delivery and patient experience during the pandemic.

Ethics Approval

All research procedures were reviewed and approved by the (deidentified) institutional review board (number 20-408; IRB) of the outpatient clinics of the (deidentified) health care system. Eligible participants were contacted using an IRB-approved study recruitment letter with instructions to contact the research team if interested in participation and informed consent obtained from them.

Recruitment

We conducted in-depth telephone interviews with pregnant women to understand their prenatal care needs, preferences, and experiences during the COVID-19 pandemic. Inclusion criteria included women who were 18 years of age or older, were English speaking, were able to provide informed consent for research participation, received obstetric care at 1 of the outpatient clinics of the (deidentified) health care system, had a viable intrauterine pregnancy, and had not experienced pregnancy loss at the time of study participation. Recruitment was structured to seek input from 2 groups of women who represent patients at different significant time points in pregnancy. One group included women in the first trimester of pregnancy to capture prenatal care needs, preferences, and experiences at the onset of the pregnancy and prenatal care delivery. A second group included women in the second trimester, whose prenatal care had been established prior to the declaration of a COVID-19 pandemic in the state of Ohio (March 2020). We conducted interviews from May to July 2020. During this time, telehealth was instituted across the health care system and, while encouraged, not required. After providing informed consent, participants took part in a semistructured telephone interview. The interviews were conducted using an interview guide that contained probes to explore pregnant patients' knowledge and perception of the COVID-19 pandemic and the pandemic's impact on their ability to access different aspects of prenatal care. The interview guide was developed in conjunction with content experts in obstetrics, health care communication, medical decision making, clinical genetics, and maternal-fetal medicine. It was piloted before use; the final version was modified based on initial participant feedback. Interviews were audio-recorded and transcribed for analysis.

Statistical Analysis

Analysis was approached as an iterative and progressive process of data immersion, coding, memoing, and theme identification, an inductive process consistent with grounded theory [18,19]. We identified content domains and categories in transcripts to

create a coding tree used to organize the data. A companion codebook was created to serve as a reference for the analysis. The coding and analysis processes were led by 2 members of the study team (authors RF and MP) using NVivo version 12 (QSR International). The research team held weekly meetings to review data coding and memoing and identify themes. Themes identified were contextualized with information about the trimester of pregnancy, gravity/parity, and previous pregnancies.

Results

Patient Characteristics

We contacted 115 (first trimester) and 139 (second trimester) patients scheduled at various outpatient clinics within this health care system. Of those, we recruited a total of 40 (15.7%)

pregnant women: 20 (50%) in their first trimester and 20 (50%) in their second trimester. The mean patient age was 32.25 (SD 4.54) years, this was the first pregnancy for 15 (37.5%) patients, 3 (7.5%) patients had symptoms of COVID-19 during this pregnancy, 5 (12.5%) were tested for COVID-19, and all results were negative (Table 1). Qualitative analysis identified 4 major themes: (1) perceptions of the benefits of telehealth during the pandemic, (2) the need for reassurance that comes from in-person clinical visits with an obstetric provider, (3) added concerns about being responsible for determining the well-being of the pregnancy at home, and (4) the impact of telehealth on patient experience with pregnancy and prenatal care (see Table 2 for additional qualitative data). No major differences in thematic identification were noted among women in the first trimester of pregnancy (study ID denoted by G1) and women in the second trimester of pregnancy (study ID denoted by G2).

Table 1. Demographics (N=40).

Demographics of participants	Participants, n (%)
Age (years)	
Non-AMA[a] (<35)	27 (67.5)
AMA (≥35)	13 (32.5)
Race	
White	34 (85)
Black	4 (10.0)
Asian	2 (5.0)
Reproductive history	
Primigradiva	15 (37.5)
Multigradiva	25 (62.5)
Trimester of pregnancy	
First trimester	20 (50.0)
Second trimester	20 (50.0)
Tested for COVID-19	
Yes	5 (12.5)
No	35 (87.5)
COVID-19 results (N=5)	
Positive	0 (0)
Negative	5 (100)
Use of telehealth (prior to interview)	
In-patient visit only	32 (80.0)
Telehealth visit	8 (20.0)

[a]AMA: advanced maternal age.

Table 2. Supplemental qualitative data.

Theme	Illustrative Quotes
Theme 1: Perceptions of the benefits of telehealth during the pandemic	• "Truthfully, for a lot of the early prenatal visits, you just sat in the waiting room for what felt like 10 hours, you peed in a cup, and you waited in the exam room for 10 hours. Then, it was a quick Doppler belly check, a quick 'How are you doing?', [a] blood pressure check, [and] 'All good. Great. We'll see you in a month.' So that's it. I can do the blood pressure at home. I could do the Doppler, theoretically. I could do weight at home." [G1, patient 13] • "Well, I think it is very cool that they're doing virtual visits. One thing that I thought was really neat is that, at my doctor's appointment, they gave me a Doppler so I can listen to the baby's heartbeat at home . . . I was very pleasantly surprised. I felt like, you know, unless there's something major going on, that a virtual visit every now and then was totally appropriate." [G1, patient 12] • "Virtual visits are a big deal . . . I was comfortable doing one. Of course, if there was something I was concerned about it, I could come in. I think it is just less enjoyable to come in because, you know, in the back of your head, you are worried about the possibility of being exposed to someone, especially if you are in a medical building." [G2, patient 4] • "I feel great about it. They sent me home with a blood pressure cuff and I want to say . . . a Doppler. I think that, with the baby and stuff, it really . . . you minimize your exposure, and you don't have to drive back and forth and sit in a waiting room. So, I feel like as long as you're healthy and everything's good, then these are great options." [G2, patient 10]
Theme 2: Reassurance that comes from in-person clinical visits with an obstetric provider	• "She [her provider] said she'd send us home with Dopplers to listen to the baby's heart at home, which is like, 'Woah. I'm supposed to sit there and try to find it?' I can't even think about that. So, it's definitely weird and different and not what I expected at all. It makes me nervous that the doctor won't be right there to do it for me, like someone who went to school for this and is trained in this." [G1, patient 3] • "I don't necessarily feel a virtual visit with my provider offers the same thing that an in-person appointment would, for example, with measuring your fundal height or making sure the heart rate is good. Unless you have the tools at home or the knowledge, you wouldn't have the ability to make sure you're on track compared to your provider doing that in person." [G1, patient 10] • "I would like more in person visits versus the telehealth, just for my own peace of mind that the pregnancy is progressing as it should." [G2, patient 12] • "I just feel like it would feel safer for me to actually see a doctor and actually have them look at me, instead of having the virtual visit. I was never worried or concerned about going to the doctor to get checked." [G1, patient 8] • "Being the first child, I just think it's nice going into the doctor especially early on when, even though you might have symptoms like you're tired or nauseous, I don't have much of a stomach yet, I don't feel the baby • kicking. So, it's nice to go in and have that reassurance that [the] baby is still growing properly and everything. So, I would prefer going in." [G1, patient 11]
Theme 3: Added concerns about the responsibility of determining the well-being of the pregnancy at home	• "The only thing I would be worried about is if something was wrong. Like, they need to make sure they hear a heartbeat or something like that having more fetal movement when I was further along." [G1, patient 4] • "I went home, and I tried to use it, and it was a couple days later and I couldn't get it to work, and I was worried it was because there was nothing there to listen to." [G1, patient 13]
Theme 4: The impact of telehealth on patient experience with pregnancy and prenatal care	• "But I think it's also important to be able to do the in-person visits, and there's some things I would imagine could get missed just by not having that in-person contact, so I see why that's important as well." [G1, patient 14] • "Right now, my regular physician has been going teledoc, so we've had to do a couple of those. The worst part about it is maybe the connection when you're standing at a laptop and there might be, like, a delay. So, that is just something you gotta take, I guess, if you're doing virtual. I wouldn't not be interested in it." [G2, patient 9] • "I don't necessarily feel a virtual visit with my provider offers the same thing that an in-person appointment would." [G1, patient 10]

Perceptions of the Benefits of Telehealth Visits During the Pandemic

For some participants, their first exposure and opportunity to use telehealth were because of the changes made in response to the pandemic. Some discussed the convenience of this approach. As described by 1 participant who had not elected for telehealth visits before the pandemic,

> I realized that the only things that were different with a virtual visit would be checking the heart rate and checking the blood pressure. Once I realized I wasn't really missing out on much and that I can check the heart rate at home by myself, made me feel better

XSL•FO

RenderX

about them [telehealth visits]. I wasn't missing out on anything. [G2, patient 2]

This finding was noted among participants in the first trimester who were beginning their prenatal care and participants in the second trimester who had already had a series of in-person visits with their provider prior to the pandemic.

Yet, for many, the benefit of telehealth was the ability to reduce the risk of exposure to COVID-19. A recurrent subtheme was concern about the "the possibility of being exposed to someone" [G2, patient 2] while presenting for an in-person prenatal care visit. For some, "virtual visits" (the term used for telehealth in this health care system) felt "safe" and "safer" than going into the office during the pandemic to avoid unnecessary potential exposures during the visit for themselves and their family. Although many participants noted the steps health care clinics and providers had taken to prevent COVID-19 exposure, there was a concern about other patients who presented for care.

> *I think that the home virtual visits are a really great option so I don't have to sit in the waiting room. Even though everyone has a cloth mask on, it's just a nice comfort level to avoid unnecessary exposures to who knows what. [. . .] I really looked forward to my prenatal visits in my other pregnancies. But, sitting in the waiting room with everybody and masks on, and seeing the front door be so busy with people, it just made me more aware of all the people that are touching surfaces and adding to the potential exposure. Because, you can't really, really trust all people.* [G1, patient 13]

For these participants, telehealth visits added a degree of safety and reassurance.

The Need for Reassurance That Comes From In-Person Clinical Visits With an Obstetric Provider

Although some participants spoke of the benefits of telehealth during the pandemic, others discussed reservations with this approach. Several participants discussed how they weighed their needs and preferences for reassurance with an in-person visit against the risk of possible COVID-19 exposure by presenting to an office or health care center.

> *It would feel safer for me to actually see a doctor and actually have them look at me instead of having the virtual visit.* [G1, patient 8]

Participants were aware of the risks of COVID-19 for themselves and their families. Yet, many also felt increased uncertainty from the pandemic and stated a preference for in-person visits to alleviate that uncertainty. As described by this participant who initiated prenatal care during the initial phases of the pandemic,

> *When I first found out I was pregnant, it was scary and stressful. When I found out, that's when the whole pandemic started. So, it was very, very stressful at first. The only thing that we were thinking of was I have a way to protect myself [. . .]. I just feel like it would feel safer for me to actually see a doctor and actually have them look at me instead of having the*

virtual visit. I was never worried or concerned about going to the doctor to get checked. [G1, patient 8]

The preference for an in-person visit was noted among women in the first and second trimester of pregnancy. Some participants in the first trimester of pregnancy preferred in-person visits to reassure them of the status of the pregnancy, particularly before there were visible or tangible changes of pregnancy. This reassurance is something that they did not feel could come from a telehealth visit. As described by this participant in the first trimester,

> *I wanted to make sure I was still pregnant even though I took like ten tests and they all said, "Yes." I just definitely wanted to make sure I still am . . . I want to make sure everything's gonna be okay.* [G1, patient 3]

This was also observed among women in the first trimester who had never been pregnant.

> *My doctor said that, maybe down the line, we could do some virtual visits. I thought, oh, that's interesting. I've never done those. I would save some time and energy on those. But at the same time, I still feel like, with my first pregnancy, I should still get to the office and have [one-on-one], face-to-face.* [G1, patient 7]

This was a theme noted across both trimesters. For example, participants in the second trimester also sought reassurance until they could experience tangible evidence of the pregnancy.

> *I still wanted to go into the doctor at least until I could feel the baby moving because I don't feel many symptoms when I'm pregnant. So, I just wanted confirmation that there was still a baby in there.* [G2, patient 21]

Added Concerns About Responsibility for Determining the Well-being of the Pregnancy

Participants also spoke of their concerns about managing their prenatal care with home monitoring and during their telehealth visit. These concerns augmented existing concerns about their health and the well-being of the pregnancy during the pandemic. One main subtheme was that although participants were provided with the instructions and tools to check the presence and status of the fetal heart tones, they noted some concern over whether they had the skills and competence to use the equipment. This concern stemmed from an underlying concern that "something would be missed" or "overlooked" if a trained health care provider did not perform the evaluation in person.

> *So, they gave me a blood pressure monitor and a Doppler. So, I guess, I am going to be in charge of doing those things, which is making me anxious.* [G1, patient 2]

Participants worried both about missing a potential health problem and about misinterpreting a result so that they would suspect a problem that did not really exist. As described by this patient,

> *So, my OB [sic] also does the virtual visits, and they said, "We'll give you a blood pressure cuff. We'll give you a Doppler. You can check your own baby's heart*

rate." And to me, I mean I know they're taking precautions, everybody . . . all the doctors are trying to see less patients and all that. But I feel like, being pregnant, I feel like that should have to be something an OB [sic] should have to physically evaluate their patient for, not me doing it myself. [. . .] I just feel like it's an important thing that the OB should be the one to physically assess the baby and me. Check my blood pressure for me because I'm not . . . I'm not trained in that. I feel like that's their job and they're the ones that are the trained professionals and they would know how to do that better than I would. [G2, patient 17]

Several participants also had a profound fear of discovering an obstetric complication or pregnancy loss while using the equipment provided for at-home monitoring. As described by this participant who was provided a Doppler to hear the fetal heart tones,

The Doppler . . . I tried using it a couple times, [and] it's still early on in the pregnancy, so it is harder to pick up on. Sometimes I'm like, "Oh gosh. Why don't I see it? Is something wrong?" [G1, patient 11]

This was also a prominent concern among participants who had experienced a pregnancy loss or complication in an earlier pregnancy. As described by this participant,

I know my next appointment will be virtual and I'll need to do my own Doppler, and that's definitely different. So, I'm not sure how I'll feel about it, especially with the scare we had [referring to a threatened miscarriage]. I'm not sure if I can't find it [heartbeat], or with me being early on, if I can't find something. It would be a huge worry on me. [G1, patient 6]

The Impact of Telehealth on Patient Experience

The meaning and significance of the experience of pregnancy were an important topic for participants. Participants sought meaningful and personal experiences with the pregnancy, something many felt that in-person access to a health care provider could offer.

If I don't go [into the office], I don't get to see the progress of my baby, and I want to be able to see that still despite the pandemic. [G1, patient 20]

There was also a sense of confronting and managing a sense of loss in the expected experience and the reality of prenatal care as it was unfolding during the pandemic.

I never imaged my first pregnancy would be like this and that I'd be talking to my doctor about a lot of things over the phone or with a mask on in our visits. [G1, patient 3]

Participants spoke of the desire to have an experience where they could align their care with their personal needs and preferences during pregnancy. Some participants noted that communication during a telehealth visit was different than their expectations for their prenatal care or than their experiences during another kind of telehealth visit. For participants, there seemed to be greater familiarity or effectiveness during an in-person interaction with the health care provider.

I prefer in-person visits because sometimes, I think of more questions to ask or I have certain concerns. So, I feel like I get better answers and support in-person. [G1, patient 5]

Interpretation of nonverbal communication, gestures, body posture, and facial expressions was also a concern for telehealth visits. As this participant explained,

I do wonder if you'll be able to catch everything that you would have [at] an in-person appointment. I know you go through your basics. You go through your basic numbers like the fetal heart rate, blood pressure, and such. [. . .] In person, with facial expressions and such, you can pick up on some subtleties. So, I do wonder if something might get lost. [G2, patient 23]

For some patients, the personal nature of pregnancy, with discussions of reproduction, children, and family, presented a sharp contrast to other types of medical visits. This participant described the uniqueness of obstetric visits and conversations compared to other health care encounters:

But it is a . . . it's a very personal thing, and I think that you kind of need to be with your doctor physically. I think it's just a different kind of care. It's a different kind of appointment, and so, I think it's a mistake to treat OB/GYN care just like any other doctor's appointment, like, 'cause, I just think it's different in so many ways. [G2, patient 18]

Discussion

Principal Findings

Our study provides important insight into the nature and scope of health care delivery challenges caused by the COVID-19 pandemic, specifically those associated with the implementation of telehealth as a formative strategy to control infection spread among health care providers, patients, and communities. Although previous studies demonstrated that telehealth is an acceptable and available way to obtain prenatal care for some pregnant patients [2,6,20], the majority of these studies were conducted prior to the pandemic and with patient populations who may have been more accepting of or able to use this approach. As a result, there is little data on the impact of broad and rapid telehealth use across diverse patient and health care provider populations in response to a pandemic.

Our study sheds light on patients' experiences and attitudes about integrating telehealth in prenatal health care delivery. Many participants in this study supported the use of telehealth in their prenatal care, citing reasons, including a sense of greater convenience, less time commitment, and less frustration from logistic steps of navigating the clinical setting compared to in-person visits; similar reasons had been noted in previous studies on the use of telehealth. In addition, they valued telehealth as a way to avoid possible exposure to COVID-19 and the potential negative consequences of infection for themselves, the pregnancy, and their family. Such factors may

help patients who may be unsure of or inexperienced with using the internet for health care to try telehealth [21-25].

However, there were several concerns about telehealth broadly and its rapid integration during the pandemic, concerns that have not been reported in earlier studies of the use of telehealth in prenatal health care delivery. A leading finding was that many participants sought additional reassurance about their health care during the pandemic, particularly the implications of COVID-19 for the pregnancy and the best approaches to managing key aspects of their pregnancy amidst the pandemic (eg, accessing testing during pregnancy, management of labor and delivery). Several participants felt that they could only achieve the level of reassurance they needed from an in-person visit. This was due, in part, to a concern that health care communication may not be as effective during an internet-based interaction compared to an in-person visit. Participants described how they viewed the use of telehealth for obstetric care compared to its use for general health care concerns. Reasons included the need for additional care and sensitivity needed for personal discussions pertaining to reproductive health, the pregnancy, parenthood, and family. For these participants, different and unfamiliar approaches to patient-provider communication also led to questions about health care quality and patient experience during their pregnancy as it was uniquely affected by the ongoing pandemic. Among this group, there was a call for health care systems to adapt telehealth approaches that would address the nuances of pregnant patients and prenatal care delivery.

Other important concerns were described. As part of telehealth implementation, participants were often provided with home monitoring equipment, such as a fetal Doppler to assess the fetal heartbeat at home. Most felt satisfied with the education provided by their health care provider about how to use this equipment. Yet, several were concerned about using this equipment at home [26,27], specifically whether they had the ability to correctly use the Doppler to hear the heart tones. This is a unique finding that calls for further investigation as other studies show patients' acceptance of using home monitoring tools [28]. Yet, a recent study demonstrated that 68% of participants were comfortable using a fetal Doppler for fetal heart tones [26], raising a question of the perspectives of those participants who did not. Further research is needed to understand how diverse patient populations may be able to implement home monitoring plans as part of telehealth, an issue that may be affected by health literacy and other key patient demographic factors.

Another concern pertained to a fear that they would be the ones to directly identify miscarriage or fetal loss by detecting the lack of a heartbeat using the fetal Doppler. Prior studies demonstrate the fears and experiences of women regarding the experience of miscarriage, particularly among those who experienced loss in a prior pregnancy [29-33]. These are significant considerations that must be addressed in telehealth implementation, both for patients experiencing pregnancy during the pandemic and for those who may use telehealth under different circumstances.

These findings raise questions about how patient experience and health care communication, key metrics to health care quality and safety, may be altered by telehealth, particularly when broadly implemented during a pandemic or other similar public health crises. Obstetric patients may have unique needs with the implementation of telehealth during the pandemic, given the myriad of uncertainties that COVID-19 has introduced to health care deliveries and communities. A model by Peahl et al [6] suggested that a 4-1-4 method may be possible for prenatal care, describing a process that combines in-person with telehealth visits that may balance the preferences and needs of patients, while also maintaining health and safety during an infectious public health threat. Thus, there is a need to develop approaches to telehealth for prenatal care delivery that align with the preferences and needs of diverse pregnant patient populations.

Limitations

Although our study sheds light on the emerging challenges with telehealth implementation, there were some limitations. The study used qualitative methods among a sample of patients at a single health care system. Our sample was limited in size and in racial and ethnic representation among diverse populations. In addition, study participants were from a health care system that could readily deploy telehealth, which may reflect participants' interest and exposure to this approach. Therefore, further research is needed to examine study findings among diverse patient populations and as the course of the pandemic continues.

Conclusion

In conclusion, although previous studies have demonstrated that telehealth is a flexible and convenient alternative for some prenatal appointments, our study suggests that there may be specific needs and concerns among the diverse patient groups using this modality during the pandemic. More research is needed to understand patients' experiences with telehealth during the pandemic and to develop approaches that are responsive to both the needs and preferences of patients and the challenges presented by public health emergencies that call for significant changes in health care delivery. This includes understanding the experiences of not only pregnant patients but also patient populations who may also have experienced the rapid conversion to telehealth in response to the pandemic.

(NGPX), and health care systems in Sweden and Saudi Arabia on topics related to public health, bioethics, and health policy. R Farrell was a consultant on the AGOG OPSA FPAR 2.0 project. This role has ended.

References

1. Zork N, Aubey J, Yates H. Conversion and optimization of telehealth in obstetric care during the COVID-19 pandemic. Semin Perinatol 2020 Oct;44(6):151300 [FREE Full text] [doi: 10.1016/j.semperi.2020.151300] [Medline: 32928561]

2. DeNicola N, Grossman D, Marko K, Sonalkar S, Butler Tobah YS, Ganju N, et al. Telehealth interventions to improve obstetric and gynecologic health outcomes: a systematic review. Obstet Gynecol 2020 Feb;135(2):371-382 [FREE Full text] [doi: 10.1097/AOG.0000000000003646] [Medline: 31977782]

3. ACOG Committee Opinion. Implementing telehealth in practice. Obstet Gynecol 2020;135(2):e73-e79. [doi: 10.1097/AOG.000000000000367]

4. Lowery C. Telehealth: A New Frontier in OB/GYN. 2018. URL: https://www.contemporaryobgyn.net/view/telehealth-new-frontier-obgyn [accessed 2022-03-28]

5. Spiegelman J, Krenitsky N, Syeda S, Sutton D, Moroz L. Rapid development and implementation of a Covid-19 telehealth clinic for obstetric patients. NEJM Catal Innov Care Deliv 2020:1-13. [doi: 10.1056/CAT.20.0170]

6. Peahl AF, Smith RD, Moniz MH. Prenatal care redesign: creating flexible maternity care models through virtual care. Am J Obstet Gynecol 2020 Sep;223(3):389.e1-389.e10 [FREE Full text] [doi: 10.1016/j.ajog.2020.05.029] [Medline: 32425200]

7. Holcomb D, Faucher M, Bouzid J, Quint-Bouzid M, Nelson D, Duryea E. Patient perspectives on audio-only virtual prenatal visits amidst the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic. Obstet Gynecol 2020;136(2):317-322. [doi: 10.1097/aog.0000000000004026]

8. Siston AM, Rasmussen SA, Honein MA, Fry AM, Seib K, Callaghan WM, Pandemic H1N1 Influenza in Pregnancy Working Group. Pandemic 2009 influenza A(H1N1) virus illness among pregnant women in the United States. JAMA 2010 Apr 21;303(15):1517-1525 [FREE Full text] [doi: 10.1001/jama.2010.479] [Medline: 20407061]

9. Silasi M, Cardenas I, Kwon J, Racicot K, Aldo P, Mor G. Viral infections during pregnancy. Am J Reprod Immunol 2015 Mar 13;73(3):199-213 [FREE Full text] [doi: 10.1111/aji.12355] [Medline: 25582523]

10. Driggers RW, Ho C, Korhonen EM, Kuivanen S, Jääskeläinen AJ, Smura T, et al. Zika virus infection with prolonged maternal viremia and fetal brain abnormalities. N Engl J Med 2016 Jun 02;374(22):2142-2151. [doi: 10.1056/nejmoa1601824]

11. Schwartz DA. An analysis of 38 pregnant women with COVID-19, their newborn infants, and maternal-fetal transmission of SARS-CoV-2: maternal coronavirus infections and pregnancy outcomes. Arch Pathol Lab Med 2020 Jul 01;144(7):799-805 [FREE Full text] [doi: 10.5858/arpa.2020-0901-SA] [Medline: 32180426]

12. Centers for Disease Control and Prevention. Data on COVID-19 during Pregnancy: Severity of Maternal Illness. URL: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/special-populations/pregnancy-data-on-covid-19.html [accessed 2020-12-07]

13. Zambrano LD, Ellington S, Strid P, Galang RR, Oduyebo T, Tong VT, CDC COVID-19 Response Pregnancy and Infant Linked Outcomes Team. Update: characteristics of symptomatic women of reproductive age with laboratory-confirmed SARS-CoV-2 infection by pregnancy status - United States, January 22-October 3, 2020. MMWR Morb Mortal Wkly Rep 2020 Nov 06;69(44):1641-1647 [FREE Full text] [doi: 10.15585/mmwr.mm6944e3] [Medline: 33151921]

14. Breslin N, Baptiste C, Gyamfi-Bannerman C, Miller R, Martinez R, Bernstein K, et al. Coronavirus disease 2019 infection among asymptomatic and symptomatic pregnant women: two weeks of confirmed presentations to an affiliated pair of New York City hospitals. Am J Obstet Gynecol MFM 2020 May;2(2):100118 [FREE Full text] [doi: 10.1016/j.ajogmf.2020.100118] [Medline: 32292903]

15. Farrell R, Michie M, Pope R. Pregnant women in trials of Covid-19: a critical time to consider ethical frameworks of inclusion in clinical trials. Ethics Hum Res 2020 Jul 20;42(4):17-23 [FREE Full text] [doi: 10.1002/eahr.500060] [Medline: 32562594]

16. Gadson A, Akpovi E, Mehta P. Exploring the social determinants of racial/ethnic disparities in prenatal care utilization and maternal outcome. Semin Perinatol 2017 Aug;41(5):308-317 [FREE Full text] [doi: 10.1053/j.semperi.2017.04.008] [Medline: 28625554]

17. Melvin SC, Wiggins C, Burse N, Thompson E, Monger M. The role of public health in COVID-19 emergency response efforts from a rural health perspective. Prev Chronic Dis 2020 Jul 23;17:E70 [FREE Full text] [doi: 10.5888/pcd17.200256] [Medline: 32701430]

18. Corbin J, Strauss A. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory (3rd ed). Thousand Oaks, CA: SAGE Publications; 2008.

19. Ryan GW, Bernard HR. Techniques to identify themes. Field Methods 2016 Jul 24;15(1):85-109 [FREE Full text] [doi: 10.1177/1525822X02239569] [Medline: 22973420]

20. Brown HL, DeNicola N. Telehealth in maternity care. Obstet Gynecol Clin North Am 2020 Sep;47(3):497-502. [doi: 10.1016/j.ogc.2020.05.003] [Medline: 32762934]

21. Fryer K, Delgado A, Foti T, Reid C, Marshall J. Implementation of obstetric telehealth during COVID-19 and beyond. Matern Child Health J 2020 Sep;24(9):1104-1110 [FREE Full text] [doi: 10.1007/s10995-020-02967-7] [Medline: 32564248]

XSL•FO

RenderX

22.   Aziz A, Zork N, Aubey JJ, Baptiste CD, D'Alton ME, Emeruwa UN, et al. Telehealth for high-risk pregnancies in the setting of the COVID-19 pandemic. Am J Perinatol 2020 Jun;37(8):800-808 [FREE Full text] [doi: 10.1055/s-0040-1712121] [Medline: 32396948]

23.   Madden N, Emeruwa UN, Friedman AM, Aubey JJ, Aziz A, Baptiste CD, et al. Telehealth uptake into prenatal care and provider attitudes during the COVID-19 pandemic in New York City: a quantitative and qualitative analysis. Am J Perinatol 2020 Aug 09;37(10):1005-1014 [FREE Full text] [doi: 10.1055/s-0040-1712939] [Medline: 32516816]

24.   Futterman I, Rosenfeld E, Toaff M, Boucher T, Golden-Espinal S, Evans K, et al. Addressing disparities in prenatal care via telehealth during COVID-19: prenatal satisfaction survey in East Harlem. Am J Perinatol 2021 Jan;38(1):88-92 [FREE Full text] [doi: 10.1055/s-0040-1718695] [Medline: 33038898]

25.   Jean-Francois B, Bailey Lash T, Dagher RK, Green Parker MC, Han SB, Lewis Johnson T. The potential for health information technology tools to reduce racial disparities in maternal morbidity and mortality. J Womens Health (Larchmt) 2021 Feb 01;30(2):274-279 [FREE Full text] [doi: 10.1089/jwh.2020.8889] [Medline: 33211604]

26.   Peahl A, Novara A, Heisler M, Dalton V, Moniz M, Smith R. Patient preferences for prenatal and postpartum care delivery: a survey of postpartum women. Obstet Gynecol 2020;135(5):A. [doi: 10.1097/aog.0000000000003731]

27.   Kobayashi H, Sado T. Satisfaction of a new telephone consultation service for prenatal and postnatal health care. J Obstet Gynaecol Res 2019 Jul 06;45(7):1376-1381. [doi: 10.1111/jog.13987] [Medline: 31060105]

28.   Butler Tobah YS, LeBlanc A, Branda ME, Inselman JW, Morris MA, Ridgeway JL, et al. Randomized comparison of a reduced-visit prenatal care model enhanced with remote monitoring. Am J Obstet Gynecol 2019 Dec;221(6):638.e1-638.e8. [doi: 10.1016/j.ajog.2019.06.034] [Medline: 31228414]

29.   Simmons RK, Singh G, Maconochie N, Doyle P, Green J. Experience of miscarriage in the UK: qualitative findings from the National Women's Health Study. Soc Sci Med 2006 Oct;63(7):1934-1946. [doi: 10.1016/j.socscimed.2006.04.024] [Medline: 16781809]

30.   Gerber-Epstein P, Leichtentritt RD, Benyamini Y. The experience of miscarriage in first pregnancy: the women's voices. Death Stud 2009 Jan 17;33(1):1-29. [doi: 10.1080/07481180802494032] [Medline: 19140290]

31.   Limbo R, Glasser J, Sundaram M. "Being sure": women's experience with inevitable miscarriage. MCN Am J Matern Child Nurs 2014;39(3):165-174. [doi: 10.1097/nmc.0000000000000027]

32.   Melender HL. Experiences of fears associated with pregnancy and childbirth: a study of 329 pregnant women. Birth 2002 Jun;29(2):101-111. [doi: 10.1046/j.1523-536x.2002.00170.x] [Medline: 12051188]

33.   Adolfsson A, Larsson PG, Wijma B, Berterö C. Guilt and emptiness: women's experiences of miscarriage. Health Care Women Int 2004 Jul;25(6):543-560. [doi: 10.1080/07399330490444821] [Medline: 15354621]

Abbreviations

IRB: institutional review board

XSL•FO
RenderX

Original Paper

A Web-Based App for Emotional Management During the COVID-19 Pandemic: Platform Development and Retrospective Analysis of its Use Throughout Two Waves of the Outbreak in Spain

Sara Guila Fidel Kinori[1,2], PhD; Gerard Carot-Sans[3,4], PhD; Andrés Cuartero[5], BSc; Damià Valero-Bover[3,4], BSc, MSc; Rosa Roma Monfa[3], BSc, MBA; Elisabet Garcia[3], BSc, MBA; Pol Pérez Sust[3,6], BSc, MSc; Jordi Blanch[3], MD; Jordi Piera-Jiménez[3,4,7], PhD; Josep Antoni Ramos-Quiroga[1,2], MD, PhD

[1]Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain

[2]CIBERSAM, Universitat Autònoma de Barcelona, Barcelona, Spain

[3]Information Systems Directorate, Servei Català de la Salut, Barcelona, Spain

[4]Digitalization for the Sustainability of the Healthcare System DS3, IDIBELL, Barcelona, Spain

[5]Sistema d'Emergències Mèdiques, Sistema de Salut de Catalunya, Barcelona, Spain

[6]Information Systems Area, Departament de Salut, Barcelona, Spain

[7]Faculty of Informatics, Multimedia and Telecommunications, Universitat Oberta de Catalunya, Barcelona, Spain

Corresponding Author:
Jordi Piera-Jiménez, PhD
Information Systems Directorate
Servei Català de la Salut
Barcelona
Spain
Phone: 34 651041515
Email: jpieraj@uoc.edu

Abstract

Background: Quarantines and nationwide lockdowns implemented for containing the spread of the COVID-19 pandemic may lead to distress and increase the frequency of anxiety and depression symptoms among the general population. During the nationwide lockdown of the first wave of the COVID-19 outbreak in Spain, we developed and launched a web-based app to promote emotional self-care in the general population and facilitate contact with health care professionals.

Objective: This study aimed to describe a web-based app and analyze its utilization pattern throughout 2 successive waves of the COVID-19 outbreak in Spain.

Methods: Our web-based app targeted all individuals aged 18 years or more and was designed by adapting the contents of a mobile app for adjuvant treatment of posttraumatic stress disorder (ie, the PTSD Coach app) to the general population and the pandemic or lockdown scenario. We retrospectively assessed the utilization pattern of the web-based app using data systematically retrieved from Google Analytics. Data were grouped into 3 time periods, defined using Joinpoint regression analysis of COVID-19 incidence in our area: first wave, between-wave period, and second wave.

Results: The resulting web-based app, named gesioemocional.cat, maintains the navigation structure of the PTSD Coach app, with three main modules: tools for emotional self-care, a self-assessment test, and professional resources for on-demand contact. The self-assessment test combines the Patient Health Questionnaire-2 and the 7-item Generalized Anxiety Disorder scale and offers professional contact in the advent of a high level of depression and anxiety; contact is prioritized in accordance with a screening questionnaire administered at the time of obtaining individual consent to be contacted. The tools for emotional self-care can be accessed either on-demand or symptom-driven. The utilization analysis showed a high number of weekly accesses during the first wave. In this period, press releases regarding critical events of the pandemic progression and government decisions on containment measures were followed by a utilization peak, irrespective of the sense (ie, positive or negative) of the information. Positive information pieces (eg, relaxation of containment measures due to a reduction of COVID-19 cases) resulted in a sharp

increase in utilization immediately after information release, followed by a successive decline in utilization. The second wave was characterized by a lower and less responsive utilization of the web-based app.

Conclusions: mHealth tools may help the general population cope with stressful conditions associated with the pandemic scenario. Future studies shall investigate the effectiveness of these tools among the general population—including individuals without diagnosed mental illnesses—and strategies to reach as many people as possible.

Introduction

In the last months, most countries worldwide have experienced uncontrolled outbreaks of the COVID-19 soon after reporting the first case. The rapid spread of its causative agent, SARS-CoV-2, has led to an unprecedented overburdening of health care systems, prompting many governments to dictate strict lockdowns for containing the outbreak.

While such indiscriminate strategies have succeeded in containing SARS-CoV-2 spread and reducing mortality [1], evidence warns on many psychological harms potentially associated with quarantines [2], particularly among older and dependent people and those with previous psychiatric pathologies [3-5]. Alongside the quarantine-specific stressors, the uncertainty associated with the pandemic scenario and impossibility of accompanying loved ones during their hospital stay or end-of-life stage has contributed to psychological overwhelming in many cases [5].

The use of mobile health (mHealth) and web-based technologies for health (eHealth) has been shown to improve access to health care services under confinement situations. In previous years, information and communication technologies (ICT) (including mHealth and eHealth solutions) for people with mental health needs have been gradually incorporated into routine care, as also for social groups not considered regular digital consumers such as older people [6]. A successful example of ICT-based solutions for mental health is the mobile app for managing posttraumatic stress disorder (PTSD; ie, the PTSD Coach app), developed by the Veterans Affairs National Centre for PTSD and the Department of Defence's National Centre for Telehealth & Technology [7]. The PTSD Coach app has been translated to many languages—including Spanish, a translation led by our team [8]—and it is currently used in our psychiatry department as an adjuvant to face-to-face therapy for PTSD.

The Catalan Health Service, which provides universal health care to a population of 7.7 million inhabitants in Northeast Spain, has a long tradition of implementing ICT solutions for promoting integrated care and increasing the efficiency and quality of health care delivery. The overloading of the health care system experienced during the COVID-19 outbreak boosted the implementation of various digital health strategies to counteract the collapse of many health resources [9]. To our knowledge, by the time of the nationwide lockdown in our country, no mHealth solutions had been developed for supporting people with mental health needs in a quarantine or lockdown context. Thus, considering the limited access of people to the health care system and the potential of ICT-based solutions to bridge these gaps, we aimed to develop a web-based app platform for supporting emotional management among the general population during the COVID-19 outbreak in our country. The existence of the PTSD Coach app provided us with an opportunity to develop such a web-based app quickly and offer it to the local population timely.

We present herein the development, main characteristics, and launch of the web-based app, named gestioemocional.cat ("emotionalmanagement" in Catalan), and a retrospective analysis of its use throughout alternate periods of lockdown and ease of social distancing measures.

Methods

Design Objectives

The web-based app was developed in response to a request by the Catalan Health System and Catalan Ministry of Health to provide Catalonia citizens (Northeast Spain) with support for coping with the emotional struggle associated with the COVID-19 pandemic and the nationwide lockdown. The web-based app was designed to assess two main objectives: (1) to provide users with tools and guidance for emotional self-assessment and care, and (2) to establish a communication channel to reach health care professionals in the advent of worsening of mental health symptoms.

The web-based app targeted all individuals aged 18 years or older with the capacity to use a mobile app or website. Spain has a high smartphone penetration rate [10], for which the web-based app had the potential to reach a significant part of the population. Nonetheless, technical requirements of the web-based app were minimized to facilitate access. Furthermore, to achieve high acceptability and on-demand use of the web-based app [11], we decided to store the least information possible and not retain the user's historical records.

Development of the Web-Based App

The government request for developing the web-based app was made on March 26, 2020. A panel of experts consisting of 2 clinical psychologists with expertise in psychological trauma and crisis intervention and 2 psychiatrists of the Catalan Health System coordinated and oversaw the web-based app development and content generation. Considering the limited time for developing the web-based app and the effectiveness of the previously developed PTSD Coach app, we used it as a

starting point to translate its contents and adapt them to a lockdown or pandemic scenario when necessary.

The panel of experts reviewed all items of the PTSD Coach app and grouped them into three categories: (1) items that could be translated into local languages without changes in their content; (2) items that required adaptation owing to cultural differences, need for broadening the clinical scope to cover anxiety and depression symptoms, or adequateness to the quarantine or pandemic scenario; and (3) items to be removed for definite inappropriateness in accordance with the intended use of the App. Owing to the urgency of the development of the web-based app, item review was based on the clinical criteria and the expertise of panel members. Of note, 2 of them had contributed to developing and translating the PTSD Coach App and had, therefore, in-depth knowledge of the rationale behind each item. The resulting contents were used to develop the web-based app using the corporate colors and design of the Catalan Health System. The authors of the PSTD Coach app kindly provided permission to adapt and translate their App.

Time Frame and Launching Strategy of the Web-Based App

The first case of COVID-19 in Spain was officially reported in late January 2020. The Spanish government halted public gatherings on March 9, 2020, and led to a state of alarm and a nationwide lockdown on March 13, 2020. The lockdown started a progressive easing on May 10, 2020, and was definitely lifted on June 20, 2020.

The web-based app gestioemocional.cat was launched by the Catalan Health Department on April 15, 2020. Information regarding the web-based app was disseminated using television and radio advertisements, which were released through the channels of the public corporation.

Utilization Analysis and Statistics

Data for utilization analysis of the web-based app were gathered from Google Analytics. The main objective of the utilization analysis was to describe the number of users throughout the study period. We used a script to systematize queries for retrieving information on daily accesses to the web-based app. For longitudinal analysis of the use of the web-based app, the number of events was transformed into a logarithmic scale and plotted against a timeline covering the period from the web-based app launch on April 15, 2020, to data set closure on December 16, 2020. Key events regarding outbreak progression (ie, relevant press releases on epidemiological information and government decisions) and the time window of each wave of the outbreak were added to the plot. The 2 waves of the COVID-19 outbreak were defined using Joinpoint regression analysis of publicly available data on COVID-19 incidence in our area. A permutation test with 0.05 confidence and 4499 permutations was used as described elsewhere [12]. Differences in the distribution of test scores on each of the priority levels in the 3 study periods (ie, first wave, second wave, and between waves) were assessed using a Pearson chi-square test and setting the significance threshold at a 2-sided α value of .05. The rest of the analyses were descriptive, and no other hypothesis tests were conducted. All analyses and plots were performed using R package [13].

Results

Platform Components

The gestioemocional.cat web-based app has the same navigation structure as the PTSD Coach app. Table 1 summarizes the changes introduced to the PTSD Coach items to adapt them to the intended use of the gestioemocional.cat web-based app; the resulting list of items was translated to the local languages (ie, Spanish and Catalan). Figure 1 outlines the navigation structure of the web-based app. The main menu provides the user with access to three sections: (1) resources for emotional management, (2) tools for self-assessment, and (3) professional resources.

Table 1. Summary of content changes for adaptation of the PTSD Coach app into gestioemociona.cat web-based app[a].

PTSD Coach app contents	Gestioemocional.cat contents	Rationale for change
Main menu		
Manage symptoms	Resources for emotional management	Wording change (no specific syndrome)
Take assessment	Self-assessment tools	—[b]
Health care resources	Health care resources	—
Learn	—	Intended for an on-demand use; no personal history recorded
Symptoms		
Remind of trauma	Painful memories	Wording adapted to a broader scope
Disconnected from people	Self-isolation	Wording adapted to a quarantine context
Disconnected from reality	Disconnected from reality	—
Sad or hopeless	Sadness or hopelessness	—
Worried or anxious	Worries	Wording adapted to a broader scope
Angry	Negative emotions	Wording adapted to a broader scope
Unable to sleep	Unable to sleep	—
Tools		
Ambient sounds	Ambient sounds	—
Body scan	Body scan	—
Change your perspective	Change your perspective	—
Contact with others	Contact with others	—
Deep breathing	Deep breathing	—
Grounding	Grounding	—
Inspiring quotes	Inspiring quotes	—
Leisure activities	Leisure: time alone	Wording adapted to a quarantine context
Mindfulness	Mindfulness	—
Muscle relaxation	Muscle relaxation	—
Observe thoughts	Observe thoughts	—
Relationship tools	Relationship tools	—
Sleep tools	Help falling asleep	—
Soothe the senses	Soothe the senses	—
Soothing images	Soothing images	—
Thought shifting	Thought shifting	—
Time out	Time out	—
RID: coping with triggers	(Removed)	In a non–posttraumatic stress disorder setting, no specific trigger is expected
Calming thoughts (user uploaded)	(Removed)	No personal section or user history was included
Schedule worry time	(Removed)	Intended for an on-demand use
Positive imagery	(Removed)	No personal section or user history was included
My feelings	(Removed)	No personal section or user history was included
Self-assessment		
3 items (evaluate your state, assessment history, and schedule self-assessment)	Evaluate your state	No personal section or user history were included
PTSD Checklist for DSM-5[c]	Patient Health Questionnaire-2	Broadening of assessment scope (anxiety and depressive symptoms expected)

PTSD Coach app contents	Gestioemocional.cat contents	Rationale for change
	Generalized Anxiety Disorder-7	Broadening of assessment scope (anxiety and depressive symptoms expected)
Professional resources		
User can choose or be redirected toward crisis resources, professional care, or choose persons in his/her social network	Exhaustive list of telephone numbers for general health and mental health support	Adapted to local resources

[a]Unchanged contents (ie, rationale for change dashed) were directly translated to the local languages (ie, Catalan and Spanish) without content changes.

[b]—: not available.

[c]DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

Figure 1. Navigation structure of the gestioemocional.cat web-based app. CAP: primary care center (for the Catalan acronym "Centre d'Atenció Primària"), GAD-7: 7-item Generalized Anxiety Disorder scale, PHQ-2: Patient Health Questionnaire-2.
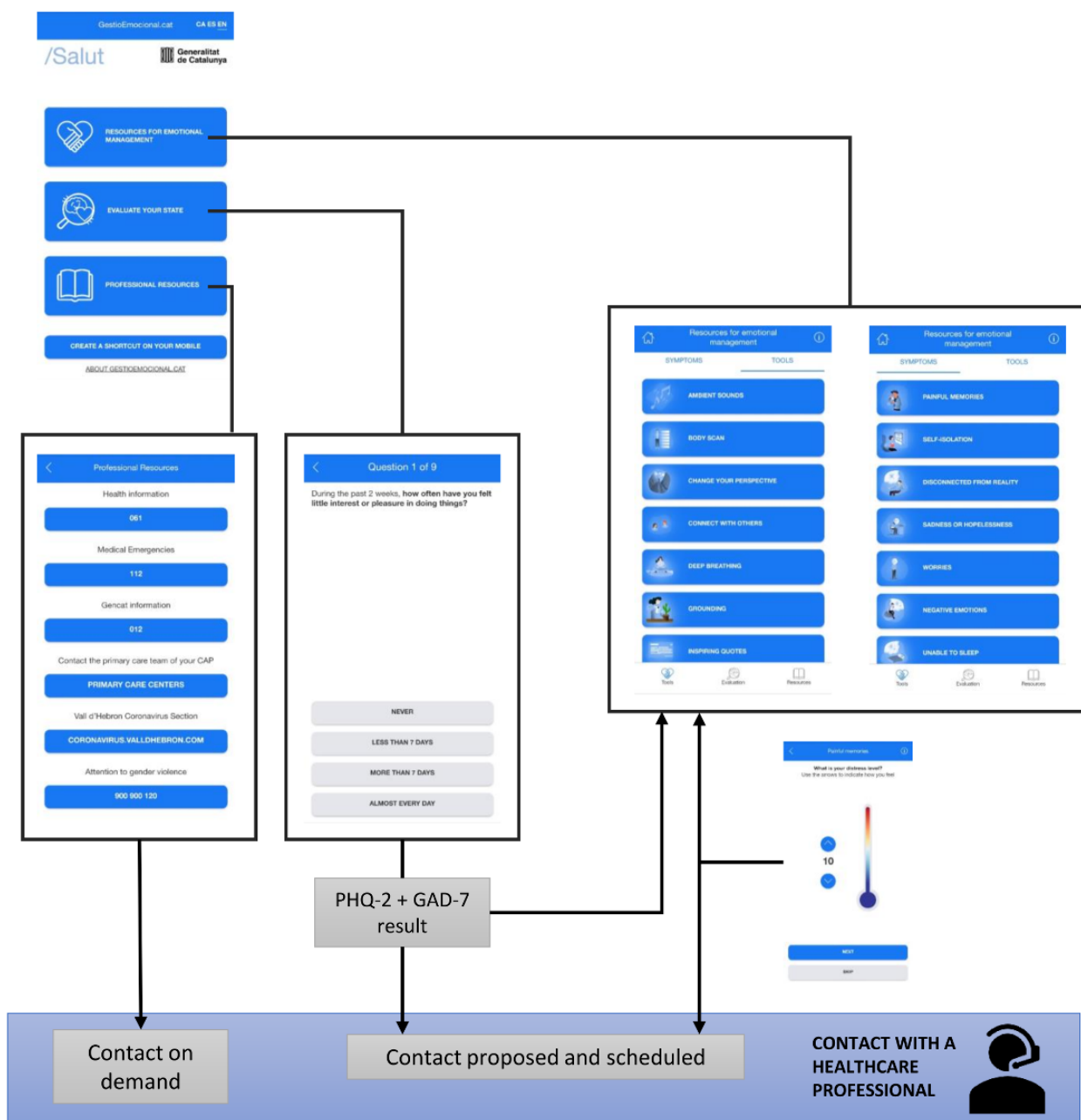
Resources for Emotional Management

The section "resources for the emotional management" can be accessed through two pathways: symptom-guided and resource type–guided (Figure 1). Symptoms include painful memories, self-isolation, disconnection from reality, sadness or hopelessness, worries, negative emotion, and sleeping disturbances. All these symptoms have been identified as very common during quarantines and lockdowns in the context of the COVID-19 pandemic [14,15]. Resource types include body scan (ie, audio-guided), change one's perspective (ie, random text messages to rationalize distress thoughts), connect with others (ie, list of ideas or tips for establishing interpersonal relationships under a quarantine context), deep breathing (ie, audio-guided technique for emotional regulation trough breathing), grounding (ie, list of tips for helping to be fully present for whatever is occurring right here and now), inspirational phrases, use of personal leisure time (ie, list of proposed activities to be performed during the lockdown), mindfulness exercises, progressive relaxation, observe thoughts (ie, audio-guided techniques to help coping with distress thoughts), relationships improvements (ie, list of tips to improve peer communication), help falling asleep (ie, list of proposals for sleeping better), calm the senses, relaxing images, thought shifting (ie, list of phrases to be repeated over 5 minutes using a countdown tool), and time-out (ie, list of proposals to change the environment when remaining quarantined).

Irrespective of the pathway chosen to access the section "resources for the emotional management," the web-based app suggests a particular resource based on the user's distress level, assessed using a single-item questionnaire (ie, what is your distress level? [use the arrows to indicate how you feel]) and rated on a 10-point scale with scores ranging 0-10. Users can select the distress level either stepwise—using up and down arrows—or using a distress thermometer with an intensity color code (Figure 1).

Self-assessment

The self-assessment section is based on two questionnaires for a mental health assessment with a Spanish version adapted and validated: the Patient Health Questionnaire-2 (PHQ-2) [16] and the 7-item Generalized Anxiety Disorder scale (GAD-7) [17]. The PHQ-2 is intended as a first-step approach to identify depressive symptoms on the basis of the frequency of depressed mood and anhedonia. While not recommended for depression diagnosis or monitoring, it is considered useful for screening for depression, which is the most common psychological conditions in clinical practice and research; individuals who

screen positive should be further evaluated with the PHQ-9 to determine whether they meet the criteria for a depressive disorder [18]. The Spanish version of the PHQ-2 has shown good reliability and validity in identifying depressive symptoms in the primary care setting [19]. The GAD-7 was developed for its use in primary care, and it has high sensitivity in identifying symptoms of the anxiety spectrum. The Spanish validation of the questionnaire has shown good reliability and validity [20]. To facilitate self-assessment, the 2 questionnaires were merged into a single questionnaire. The combination of the PHQ-2 and GAD-7 for a joint assessment of anxiety and depression has shown high internal reliability and strong convergent and construct validity when assessing its association with other mental health, quality of life, and disability measures [21]. All items of the combined questionnaire refer to the frequency of symptoms within the past 2 weeks and are rated on a 4-point scale where 0="never," 1="less than seven days," 2="more than seven days," and 3="almost every day."

The results of the 2 questionnaires are combined to provide 4 priority levels of care needs (Figure 2). Users with low scores in the PHQ-2 and GAD-7 (ie, suggestive of mild symptoms of anxiety and depression) are encouraged to the routine use of self-management and self-assessment tools to either maintain (level 4) or improve (level 3) their emotional health status. Conversely, users with high scores on the PHQ-2 and GAD-7 (ie, suggestive of severe anxiety and depressive symptoms) are offered telephone contact with professionals of the emergency medical service. Users who agree to be contacted by a health care professional are redirected to a consent form, which collects minimum personal data. Contacts are prioritized on the basis of the combined result of the self-assessment questionnaire (ie, 1 or 2) and a screening form asking about 6 extreme mental health conditions: suicidal ideation, loss of a loved one due to COVID-19, family member admitted to hospital owing to COVID-19, living with someone at high risk of severe illness or with COVID-19 symptoms, current psychiatric treatment, and previous psychiatric treatment. Affirmative responses to these questions were assigned the highest priority, irrespective of the self-assessment questionnaire results (Figure 2).

Scheduled telephone calls are performed by the same professionals attending calls from people who reach the emergency medical service seeking mental health support. Once the phone contact has been established, health care professionals perform an independent assessment, irrespective of the contact source (ie, direct call to emergency medical services or scheduled call through the web-based app).

Figure 2. Prioritization algorithm for proactive contact with health care professionals on the basis of the assessment of mental health status. GAD-7: 7-item Generalized Anxiety Disorder scale, PHQ-2: Patient Health Questionnaire-2.

Healthcare Resources

The section "healthcare resources" allows the user to proactively contact community resources, including general health information, emergency medical services, general government information office, the user's primary care center, and the office for women under gender violence. Additional resources include the Galatea Foundation, aimed at providing support to health care professionals, the Official College of Psychologists, which can respond to population demands and transfer cases to public health care professionals, and the Vall d'Hebron Coronavirus Section, a web-based platform available 24 hours a day, 7 days a week for attending concerns related to emotional and social impact of the COVID-19 pandemic.

Utilization

Between April 15, 2020 (ie, the launch date), and December 16, 2020, the web-based app reported 582,826 accesses. Figure 3 shows the longitudinal analysis of the number of users and critical events of pandemic progress. During the final period of the nationwide lockdown, the daily number of users peaked concomitantly with key news or government actions released by the press. Most of the press releases announcing changes on containment measures were immediately followed by a utilization peak, irrespective of the sense of these changes (ie, increasing or easing restrictions). Immediately after announcing the easing of restrictions (eg, the start of the de-escalation plan), the daily number of users remarkably decreased. The decrease was particularly abrupt when Barcelona, the most populated area, entered the first phase of the de-escalation plan.

Figure 3. Longitudinal analysis of the number of accesses to the web-based app (logarithmic scale). Each time point shows the cumulative number of users in the 5 previous days. Red dotted lines show press releases of key events and government decisions regarding the COVID-19 outbreak. A: April 17, 2020—the Catalan Ministry of Health reports and posts the highest RT-PCR–positive and hospitalization rates. B: April 28, 2020—the Spanish government announces a de-escalation plan, to be deployed in 4 phases. Catalonia exceeds 10,000 deaths due to COVID-19. C: May 5, 2020—the government announces an upcoming partial lockdown. D: May 25, 2020, Barcelona—the greatest area in Catalonia, enters de-escalation phase 1. E: June 21, 2020—the whole country officially starts the "new normal" state. F: October 16, 2020—the government dictates a temporal lockdown of bars and restaurants. G: October 25, 2020—the government dictates nationwide curfew from 10 PM to 6 AM. H: October 30, 2020—the government dictates a permanent closure of area boundaries and weekend closure of city boundaries. The time window of each period (ie, first wave, between waves, and second wave) was defined using a Joinpoint regression analysis. RT-PCR: reverse transcription–polymerase chain reaction.
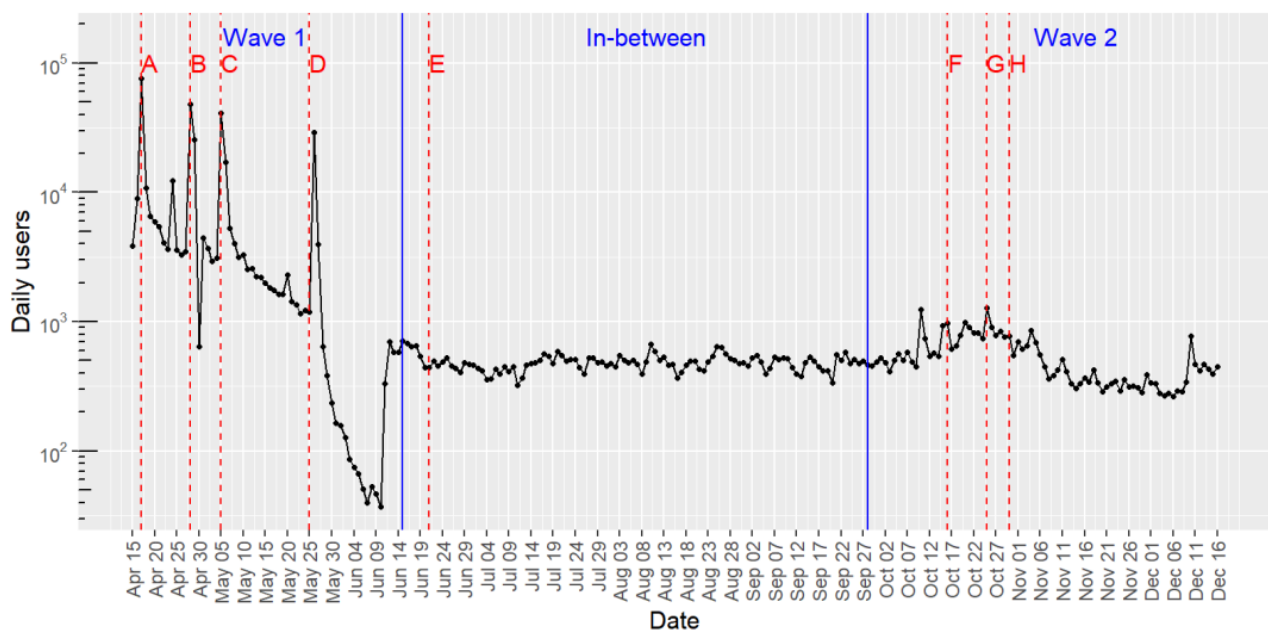
The number of users during the second wave of the outbreak was persistently lower than that in the first wave. During this period, use of the web-based app was less sensitive to the announcement of government measures.

Progression of Self-assessment

During the study period, the web-based app recorded 142,337 self-assessment tests: 124,087 (87.2%) during the first wave, 8510 (6.0%) during the period between waves, and 9740 (6.8%) during the second wave. The chi-square test revealed significant differences in the distribution of percentages across the priority levels (obtained from the self-assessment module) between the analysis periods (Figure 4). The percentage of tests scoring for priority levels 1 or 2 (associated with more severe symptoms of anxiety and depression) was higher in the second wave and between-wave periods. Overall, the offer of professional help was accepted in 8588 cases (11.4% of all cases with priority levels 1 or 2 and, therefore, triggering the message for an offer of professional help).

Figure 4. Distribution of the results of the self-assessment test across priority levels (ie, risk group) in each of the analysis periods: first wave (n=124,087), between waves (n=8510), and second wave (n=9740). The 3 periods were defined using a Joinpoint regression analysis. The Pearson chi-square test for percentage distribution was significant at $P<.001$ ($\chi^2_6=1349.1$).
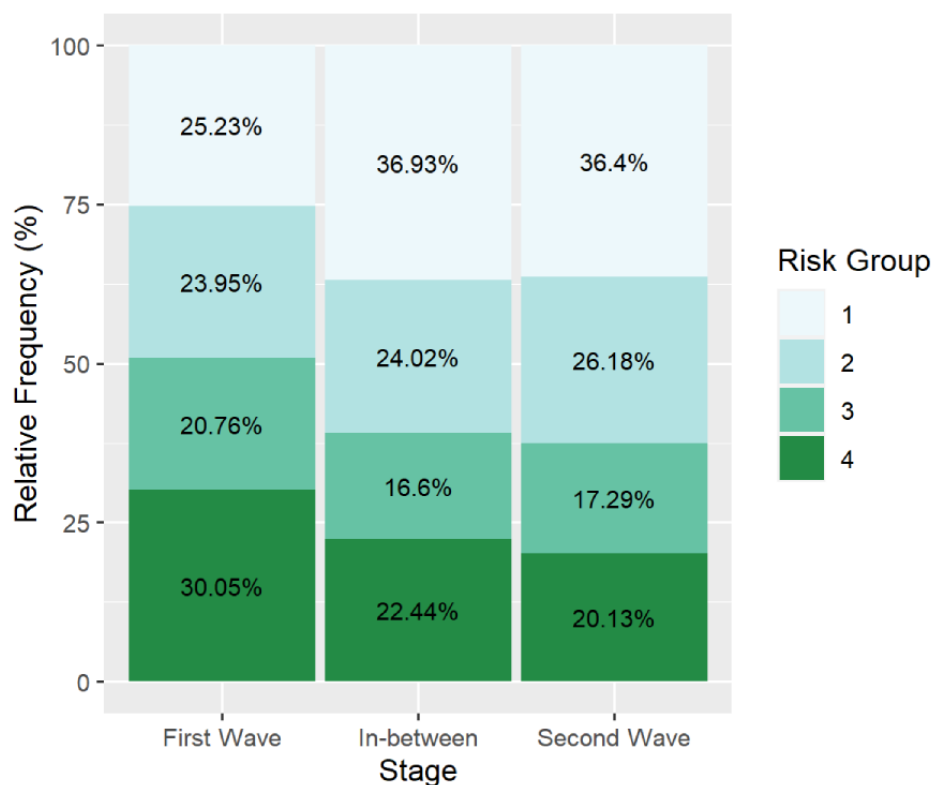
Discussion

Principal Findings

In this study, we describe a web-based app aimed at providing the general population with self-care and self-assessment tools for emotional health during the lockdown associated with the COVID-19 pandemic and establishing a new communication channel between the population and health care professionals. In addition, to reporting the characteristics of the web-based app, we analyzed the utilization pattern of the web-based app during the first 10 months of the COVID-19 outbreak in Spain. The frequency of use of the web-based app was particularly high during the first wave of the outbreak in Spain and progressively declined until reaching a baseline activity below 1000 weekly accesses after halting the national lockdown. During the first wave, the number of accesses peaked after press releases associated with the pandemic progression or changes in the containment measures dictated by the government. Interestingly, positive news (eg, relaxation of social distancing measures) were immediately followed by a sharp increase in utilization and a successive decline. The between-wave and second wave periods were characterized by a less responsive pattern of web-based app users.

The self-assessment module of the web-based app, designed to work as a sentinel for detecting severe mental health symptoms and proactively offer direct contact with health care professionals, was mostly used during the first wave. However, the scores of the self-assessment questionnaire were higher after the first wave, suggesting a higher proportion of people with

mental health care during the second wave and between-wave periods.

Limitations

Development and launching of the web-based app were strongly influenced by the urgency and overloading of the health care system, which challenged the design of an adequate assessment strategy of the platform. We chose the PTSD Coach app because of the symptom overlapping between PTSD and those reported elsewhere for lockdown situations, particularly regarding anxiety and depression. One of the advantages of the PTSD Coach app was the body of evidence on its effectiveness for self-management of mental health needs. However, owing to the shift in the target population (ie, from people diagnosed with PTSD to the general population), effectiveness of our web-based app cannot be assumed and should be confirmed in future studies. Of note, the complicated scenario in which the web-based app should be assessed will strongly challenge the trial conduct.

Another important limitation of our retrospective analysis was the limited data stored, which precluded more exhaustive assessments of behavior patterns among users (eg, to discriminate between entries of new users and regular access of the same user). However, because of the negative correlation that often exists between App use or demand and permissions for accessing private data [11], we were concerned that collecting too much data from users might be a barrier to web-based app utilization. Of note, the government announcement of launching a COVID-19 tracer app raised many privacy concerns among the population and was strongly

criticized. In light of this general perception, we prioritized the acceptability of the web-based app over the capacity of collecting personal information. As a result, our database for the retrospective analysis lacked important information that might have helped to understand behavior patterns among users. One of the most remarkable consequences of this limitation is the lack of data at a user level, which precludes discriminating between entries of new users and regular access of the same user. More importantly, the lack of historical records of users prevents us from analyzing improvement or worsening of mental health state. Finally, we could not analyze the usage of the web-based app in accordance with socioeconomic status, which has been highlighted as a source of inequality in other aspects associated with access to COVID-19 resources [22,23]. Nevertheless, while all these limitations compromise analytical approaches of the web-based app, the absence of personal questions and no need for signing in are likely to promote the platform use, in line with the overarching goal of the project, which was to help the population during the stressful circumstances of the lockdown.

Comparison With Prior Work

To our knowledge, by the time of developing the web-based app, no other mHealth solutions for promoting emotional self-care in the general population during a lockdown or quarantine situation had been launched or published. The closest example of an mHealth tool that matched our aim was the PTSD Coach app, which was innovative in the sense that users can access tools through two alternative pathways: on-demand and symptom-driven [7]. Although the PTSD Coach app was designed as adjuvant to face-to-face therapy for PTSD, its authors suggested that it could also help people not receiving psychological treatment [24,25]. Further experiences also showed good usability and ease of cultural adaptation of the App [8], encouraging us to adapt its contents to a lockdown and pandemic scenario in our country.

Although we cannot directly compare the utilization patterns of our web-based app with other mHealth or eHealth solutions for lockdown contexts, the trends observed in our analysis are consistent with information reported elsewhere. For instance, the increased consumption of alcohol and drug abuse reported during this period in our area [15] could likely be attributed to increased anxiety symptoms, grief processes, and difficulties associated with intensive cohabitation [26,27]. Similarly, other countries have reported increased demand for attention from people with mental health needs during nationwide lockdowns associated with the COVID-19 pandemic [28,29].

The design of our retrospective analysis precludes drawing strong conclusions regarding the reasons underlying the observed trend toward a less responsive usage of the web-based app during the second wave of the outbreak. However, this finding is consistent with the "pandemic fatigue" phenomenon highlighted by the World Health Organization and defined as an expected and natural response to a prolonged public health crisis [30]. This type of reaction might correspond to M Seligman's Paradigm on learned helplessness, which describes a passive behavior exhibited by a subject after enduring repeated aversive stimuli beyond their control. It is worth noticing, however, our finding that self-assessment tests more often scored priority levels 1 or 2 during the between-wave and second wave periods suggests that people with higher mental health needs were more likely to retain the usage of the web-based app after the first wave.

Finally, the fact that 8588 self-assessment tests ended up with a scheduled telephone call with health care professionals for mental health support indicates that the web-based app has emerged as an additional communication channel for seeking professional help. The number of telephone contacts represents a low percentage (11.4%) of the total situations in which professional support was offered. However, this finding is consistent with current evidence, which suggests that most people with common mental health problems do not seek professional help [31,32].

Conclusions

In the first year of the COVID-19 pandemic, various authors have emphasized the need to promote original and creative ways to help people cope with the new and stressful conditions during the pandemic [33,34]. We did not want to miss the opportunity to adapt mHealth tools with proven efficacy in self-management of stressful situations to the pandemic scenario. Future studies shall investigate the extent to which this tool helps people cope with the pandemic scenario and address strategies for making as many people as possible aware of this resource.

Conflicts of Interest

JB was on the speakers' bureau and/or acted as consultant for Janssen-Cilag, GILEAD, MSD, ViiV, Menarini in the last 3 years. He also received travel awards (transport + hotel) from Lundbeck, Janssen-Cilag, GILEAD, MSD, and Menarini for taking part in psychiatric meetings. All the other authors confirm that they have no conflicts of interests related to the manuscript.

References

1. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, Imperial College COVID-19 Response Team, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature 2020 Aug;584(7820):257-261. [doi: 10.1038/s41586-020-2405-7] [Medline: 32512579]

2. Mohamad Marzuki MF, Yaacob NA, Bin Yaacob NM, Abu Hassan MR, Ahmad SB. Usable Mobile App for Community Education on Colorectal Cancer: Development Process and Usability Study. JMIR Hum Factors 2019 Apr 16;6(2):e12103 [FREE Full text] [doi: 10.2196/12103] [Medline: 30990454]

3. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]

4. Yang Y, Li W, Zhang Q, Zhang L, Cheung T, Xiang Y. Mental health services for older adults in China during the COVID-19 outbreak. Lancet Psychiatry 2020 Apr;7(4):e19 [FREE Full text] [doi: 10.1016/S2215-0366(20)30079-1] [Medline: 32085843]

5. Arango C. Lessons Learned From the Coronavirus Health Crisis in Madrid, Spain: How COVID-19 Has Changed Our Lives in the Last 2 Weeks. Biol Psychiatry 2020 Oct 01;88(7):e33-e34 [FREE Full text] [doi: 10.1016/j.biopsych.2020.04.003] [Medline: 32381276]

6. Schueller SM, Neary M, O'Loughlin K, Adkins EC. Discovery of and Interest in Health Apps Among Those With Mental Health Needs: Survey and Focus Group Study. J Med Internet Res 2018 Jun 11;20(6):e10141 [FREE Full text] [doi: 10.2196/10141] [Medline: 29891468]

7. Kuhn E, Greene C, Hoffman J, Nguyen T, Wald L, Schmidt J, et al. Preliminary evaluation of PTSD Coach, a smartphone app for post-traumatic stress symptoms. Mil Med 2014 Jan;179(1):12-18. [doi: 10.7205/MILMED-D-13-00271] [Medline: 24402979]

8. Kuhn E, van der Meer C, Owen JE, Hoffman JE, Cash R, Carrese P, et al. PTSD Coach around the world. Mhealth 2018;4:15 [FREE Full text] [doi: 10.21037/mhealth.2018.05.01] [Medline: 29963560]

9. Pérez Sust P, Solans O, Fajardo JC, Medina Peralta M, Rodenas P, Gabaldà J, et al. Turning the Crisis Into an Opportunity: Digital Health Strategies Deployed During the COVID-19 Outbreak. JMIR Public Health Surveill 2020 May 04;6(2):e19106 [FREE Full text] [doi: 10.2196/19106] [Medline: 32339998]

10. De-Sola J, Rubio G, Talledo H, Pistoni L, Van Riesen H, Rodríguez de Fonseca F. Cell Phone Use Habits Among the Spanish Population: Contribution of Applications to Problematic Use. Front Psychiatry 2019;10:883 [FREE Full text] [doi: 10.3389/fpsyt.2019.00883] [Medline: 31920744]

11. Kummer M, Schulte P. When Private Information Settles the Bill: Money and Privacy in Google's Market for Smartphone Applications. Management Science 2019 Aug;65(8):3470-3494. [doi: 10.1287/mnsc.2018.3132]

12. Kim H. Correction: Kim H-J, Fay MP, Feuer EJ, Midthune DN, "Permutation tests for joinpoint regression with applications to cancer rates". Statistics in Medicine 2000 19:335-351. Statist Med 2001 Feb 28;20(4):655. [doi: 10.1002/sim.811]

13. The R Project for Statistical Computing. The R Foundation. URL: https://www.r-project.org [accessed 2020-05-25]

14. Xiong J, Lipsitz O, Nasri F, Lui LM, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: A systematic review. J Affect Disord 2020 Dec 01;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]

15. García-Álvarez L, de la Fuente-Tomás L, García-Portilla MP, Sáiz PA, Lacasa CM, Dal Santo F, et al. Early psychological impact of the 2019 coronavirus disease (COVID-19) pandemic and lockdown in a large Spanish sample. J Glob Health 2020 Dec;10(2):020505 [FREE Full text] [doi: 10.7189/jogh.10.020505] [Medline: 33110588]

16. Kroenke K, Spitzer R, Williams J. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]

17. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

18. American Psychological Association. URL: https://www.apa.org/pi/about/publications/caregivers/practice-settings/assessment/tools/patient-health [accessed 2021-02-20]

19. Reuland DS, Cherrington A, Watkins GS, Bradford DW, Blanco RA, Gaynes BN. Diagnostic accuracy of Spanish language depression-screening instruments. Ann Fam Med 2009;7(5):455-462 [FREE Full text] [doi: 10.1370/afm.981] [Medline: 19752474]

20. García-Campayo J, Zamorano E, Ruiz M, Pardo A, Pérez-Páramo M, López-Gómez V, et al. P-149 - Psychometric validation of the spanish version of the GAD-2 scale for screening generalized anxiety disorder. European Psychiatry 2012 Jan;27:1. [doi: 10.1016/s0924-9338(12)74316-x]

21. Kroenke K, Wu J, Yu Z, Bair MJ, Kean J, Stump T, et al. Patient Health Questionnaire Anxiety and Depression Scale: Initial Validation in Three Clinical Trials. Psychosom Med 2016;78(6):716-727 [FREE Full text] [doi: 10.1097/PSY.0000000000000322] [Medline: 27187854]

22. Amengual-Moreno M, Calafat-Caules M, Carot A, Rosa Correia AR, Río-Bergé C, Rovira Plujà J, et al. [Social determinants of the incidence of Covid-19 in Barcelona: a preliminary ecological study using public data.]. Rev Esp Salud Publica 2020 Sep 16;94:e202009101 [FREE Full text] [Medline: 32935664]

XSL•FO

RenderX

23. Baena-Díez JM, Barroso M, Cordeiro-Coelho S, Díaz JL, Grau M. Impact of COVID-19 outbreak by income: hitting hardest the most deprived. J Public Health (Oxf) 2020 Nov 23;42(4):698-703 [FREE Full text] [doi: 10.1093/pubmed/fdaa136] [Medline: 32776102]

24. Kuhn E, Kanuri N, Hoffman JE, Garvert DW, Ruzek JI, Taylor CB. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. J Consult Clin Psychol 2017 Mar;85(3):267-273. [doi: 10.1037/ccp0000163] [Medline: 28221061]

25. Ruzek JI, Kuhn E, Jaworski BK, Owen JE, Ramsey KM. Mobile mental health interventions following war and disaster. Mhealth 2016;2:37 [FREE Full text] [doi: 10.21037/mhealth.2016.08.06] [Medline: 28293610]

26. Ben-Zur H, Zeidner M. Threat to life and risk-taking behaviors: a review of empirical findings and explanatory models. Pers Soc Psychol Rev 2009 May;13(2):109-128. [doi: 10.1177/1088868308330104] [Medline: 19193927]

27. Boscarino JA, Adams RE, Galea S. Alcohol use in New York after the terrorist attacks: a study of the effects of psychological trauma on drinking behavior. Addict Behav 2006 Apr;31(4):606-621 [FREE Full text] [doi: 10.1016/j.addbeh.2005.05.035] [Medline: 15982827]

28. Liu S, Yang L, Zhang C, Xiang Y, Liu Z, Hu S, et al. Online mental health services in China during the COVID-19 outbreak. Lancet Psychiatry 2020 Apr;7(4):e17-e18 [FREE Full text] [doi: 10.1016/S2215-0366(20)30077-8] [Medline: 32085841]

29. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]

30. Pandemic fatigue: Reinvigorating the public to prevent COVID-19. World Health Organization. URL: https://apps.who.int/iris/bitstream/handle/10665/335820/WHO-EURO-2020-1160-40906-55390-eng.pdf [accessed 2021-01-05]

31. Oliver MI, Pearson N, Coe N, Gunnell D. Help-seeking behaviour in men and women with common mental health problems: cross-sectional study. Br J Psychiatry 2005 Apr;186:297-301. [doi: 10.1192/bjp.186.4.297] [Medline: 15802685]

32. Bayer JK, Peay MY. Predicting intentions to seek help from professional mental health services. Aust N Z J Psychiatry 1997 Aug;31(4):504-513. [doi: 10.3109/00048679709065072] [Medline: 9272260]

33. Inkster B, O'Brien R, Selby E, Joshi S, Subramanian V, Kadaba M, et al. Digital Health Management During and Beyond the COVID-19 Pandemic: Opportunities, Barriers, and Recommendations. JMIR Ment Health 2020 Jul 06;7(7):e19246 [FREE Full text] [doi: 10.2196/19246] [Medline: 32484783]

34. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]

Abbreviations

GAD-7: 7-item Generalized Anxiety Disorder scale
ICT: information and communication technologies
mHealth: mobile health
PHQ-2: Patient Health Questionnaire-2
PTSD: posttraumatic stress disorder

XSL•FO
RenderX
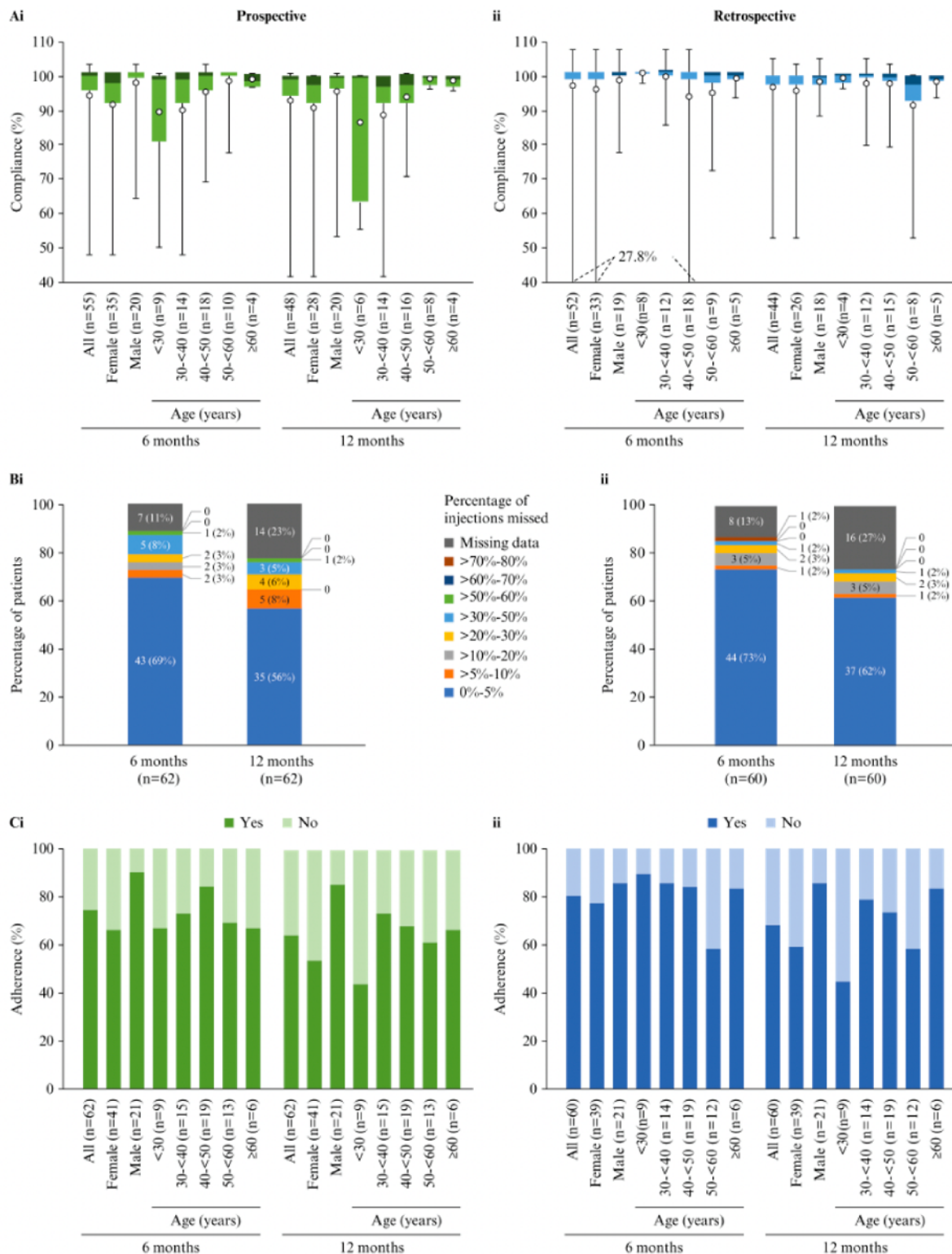
XSL•FO

RenderX

Corrigenda and Addenda

Correction: Ascertaining Medication Use and Patient-Reported Outcomes via an App and Exploring Gamification in Patients With Multiple Sclerosis Treated With Interferon β-1b: Observational Study

Volker Limmroth[1], MD; Kirsten Bayer-Gersmann[2], BEng; Christian Mueller[3], PhD; Markus Schürks[3], MSc, MD

[1]Clinic for Neurology and Palliative Medicine, Municipal Hospital Köln-Merheim, Cologne, Germany

[2]Institut Dr. Schauerte, Munich, Germany

[3]Bayer Vital GmbH, Leverkusen, Germany

Corresponding Author:
Markus Schürks, MSc, MD
Bayer Vital GmbH
Medical Affairs, Neurology, Immunology & Ophthalmology
Leverkusen, 51368
Germany
Phone: 49 214 30 51455
Email: markus.schuerks@bayer.com

Related Article:

Correction of: https://formative.jmir.org/2022/3/e31972

In "Ascertaining Medication Use and Patient-Reported Outcomes via an App and Exploring Gamification in Patients With Multiple Sclerosis Treated With Interferon β-1b: Observational Study" (JMIR Form Res 2022;6(3):e31972) the authors noted one error.

In the originally published article, Figure 2 appeared incorrectly (as shown in the Multimedia Appendix 1).

The correct Figure 2 is provided below.

The correction will appear in the online version of the paper on the JMIR Publications website on March 17, 2022, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Figure 2. (A) Compliance overall and by gender and age group, (B) patients categorized by percentage of interferon β-1b injections missed (based on an expected frequency of 1 injection every other day), and (C) adherence overall and by gender and age group, analyzed (i) prospectively and (ii) retrospectively. Compliance at 6 and 12 months was assessed in patients with ≥6 and ≥12 months of injection-related data, respectively. In the box and whisker plots, the colored bars indicate median and IQR, the whiskers indicate minimum and maximum values, and the white circles indicate the mean.

Multimedia Appendix 1
Originally published Figure 2.
[PNG File , 95 KB - formative_v6i3e38002_app1.png]

XSL•FO

RenderX